

-199-

orf75-1	ATLADMALFPERRMLMAREITKTFETFLSTGVEIQTALSADGNQSRGEMVLVLYPAQD 190 200 210 220 230 240
5 orf75a.pep	EKHGLSEGAQNIMKILTAELPTKQAAELAAKITPGEKGKKYLADLALSWKNNX 250 260 270 280 290
orf75-1	EKHGLSEGAQNIMKILTAELPTKQAAELAAKITPGEKGKKYLADLALSWKNNX 250 260 270 280 290

10 Homology with a predicted ORF from *N. gonorrhoeae*

ORF75 shows 93.2% identity over a 292aa overlap with a predicted ORF (ORF75.ng) from *N. gonorrhoeae*:

	orf75.pep	MFVFTAQXFMQKHLQKASDSDSVVGTLVYVATPIGNLADITLRLAVALQKA----AEDETR	56
15	orf75ng	MSVFTAQFFMPQKHLQKASDSDSVVGTLVYVATPIGNLADITLRLAVALQKAIDICAEDETR	60
	orf75.pep	VTAQLLSAYGIQGKLVSVEHNERQNMADKIVGYLSDGMVVAQVSDAGTPAVCDPGAKLAR	116
	orf75ng	VTAQLLSAYGIQGKLVSVEHNERQNMADKIVGYLSDGLVVAQVSDAGTPAVCDPGAKLAR	120
20	orf75.pep	RVREAGFKVVPPVGAXAVMAALSVAGVEGSGDFYFNGFVPPKGERRKLFAKWVRAAFPIV	176
	orf75ng	RVREAGFKVVPPVGASAVMAALSVAGVAESDPYFNGFVPPKGERRKLFAKWVRAAFPIV	180
25	orf75.pep	MFETPHRIGAALADMAELPERRMLAREITKTFETFLSGTVGEIQTALSADGQSRGEM	236
	orf75ng	MFETPHRIGATLADMAELPERRMLAREITKTFETFLSGTVGEIQTALSADGQSRGEM	240
30	orf75.pep	VLVLYPAQDEKHEGLSESQAQNIMKILTAELPTKQAAELAAKIKTGEGKKALYD	288
	orf75ng	VLVLYPAQDEKHEGLSESQAQNIMKILAAELPTKQAAELAAKIKTGEGKKALYDLALSWNK	300

An ORF75ng nucleotide sequence <SEQ ID 291> was predicted to encode a protein having amino acid sequence <SEQ ID 292>;

1 MSVFQTAFFM FKHLQKASD SVVGGILYVV ATPIGNGLADI TLRLAVLQK
35 51 ADIICAEADTR VAQLLSAYQ IKGRLVSVRE HNERQMADKV IGFSLGDLVQ
101 AQSVDGATPA VCDPQAKLAR RVRERAGFVV PVFGASVAMLA VLSVGVAES
151 DFTYFNGVFP KSGERKLKKA KWRAFGRPPV MFETPRHIGA TLADMALBF
201 EERRMLAREI TKTTFITLFG TVGETQITALA ADNGRSQEGR VLVLVPAQE
251 KHEGLSESAQ NAMKILAAELA PTKQAAELAA KITGECKKAL YDLALSWKNK
40 301 *

After further analysis, the following gonococcal DNA sequence <SEQ ID 293> was identified:

45	1	ATGTTTCAGA	AACACTTGTCA	GAAAGCCTCC	GACAGCGTCG	TGGAGGGAC
	51	ATTATTAACCTG	TTGCCCGAC	CCATCGGAA	TTTGGCAGAC	ATTTCCCTTG
	101	GCGCTTGGG	GGTATTGCAAA	AAAGCCGACA	TATCTTGG	CGAAAGAACACG
	151	GGCGTTTACTC	CGGCAGCTT	GACCGCTGAT	GCCATCTCAG	CGAGGTGGT
	201	CAGTGTGTCG	GRACACACAG	ACGGCGCAT	GGCGGACGAC	GAATACCGGT
	251	TCTCTTTCAGA	GGCGCTGCG	GTGGCCGAGG	TTTCCGATGTC	GGGTACCGCC
	301	GGCGTGTGCG	ACCCGGCGGC	GAACATCTCC	GGCCGGTGC	GGCGAGCAGG
	351	GTTCGAAGTC	GTTCCTCCG	TGGCGGCAAG	CGCGGTAAATG	GGGGGTGATG
50	401	GTGTGGCGCG	TGTGGCGGA	TCCGATTTTT	ATTTCACCGG	TTTTTGACCC
	451	CGGAATTCGG	GGGAACGTAG	GAARATTGTTT	GCCAAATGGG	TGGCGGCGGC
	501	ATTTCTCTGC	GTTCATCTGG	AAACGGCGCA	CGGAATCGGG	GAACCTGGT
	551	CGGATATGG	GGAAATGTTG	CCCGAGACCC	GTCTGTATGT	GGCGGGGCAA
	601	ATCACGAAA	GGTTTGAAG	GTCTTGTAGC	GGCACCGTTG	GGGAAATTAC
55	651	GACGCGCTAC	GGCGGGCGAC	GCACCAANTC	GGCGGGCGAC	ATGGTGTGTTG
	701	TGCTGTATTC	GGCCGCAATG	GAAAATACACCS	AAAGCTTGTG	CGAGGTGGCGC
	751	CAAATATCGA	TGAAATACCT	TTGGGGCAGG	CTGGCGACCA	GGGGGGGGCG
	801	GGAGCTTGTCC	GGGAAATGTT	CAGGTGAGGG	AAAAAAGGT	TTTGTACGATT
	851	TGGCACGCTTC	GTGGGAAACAC	ATAATGA		

60 This corresponds to the amino acid sequence <SEQ ID 294; ORF75ng-1>;

-200-

1 MFQKHLQKAS DSVVGGTLYKV VATPIGNLAD ITILRALAVLQ KADIICAEDT
 51 RVTAAQLLSAY GIQGRLVSR EHNERGMADK VIGFLSDGLV VAQVSDAGTP
 101 AVCDPGAKLA RRVRREAGFKVY VPVVGASAVM AALSVAGVWE SDFYFNGFVP
 151 PKSGERKLPE ARKWRAAFFV VMFETPHRIG ATTLMALIF FERRILMLARE
 201 ITKTTEFTLS GTVEIQTAL AAADGNQSRGE MVLVLYPAQD EKHEGLSESA
 251 QNAMKILAAE LPTKQAEELA AKITGEGKRA LYDLALSWKN K*

ORF75ng-1 and ORF75-1 show 96.2% identity in 291 aa overlap:

	10	20	30	40	50	60
10	orf75-1.pep	MFQKHLQKAS DSVVGGTLYKV VATPIGNLAD ITILRALAVLQ KADIICAEDT				
	orf75ng-1	MFQKHLQKAS DSVVGGTLYKV VATPIGNLAD ITILRALAVLQ KADIICAEDT				
		10	20	30	40	50
15	orf75-1.pep	70 80 90 100 110 120				
	orf75ng-1	GIQGRLVSVREHNERQMADKIVGYLSQVSDAGTPAVCDPGAKLARRVREAGFKV				
		70	80	90	100	110
20	orf75-1.pep	130 140 150 160 170 180				
	orf75ng-1	VPVVGASAVMAALSVAGVEGSDFYFNGFVPPKSGERKLFAKWRRAAFFPVMFETPHRIG				
		130	140	150	160	170
25	orf75-1.pep	190 200 210 220 230 240				
	orf75ng-1	ATLADMALFPERRMLAREITKTETFLSGTVGEIQTALSADGNQSRGENMVLVLYPAQD				
		190	200	210	220	230
30	orf75-1.pep	250 260 270 280 290				
	orf75ng-1	EKHEGLSESAQNMILAAELPTKQAAEELAAKITGEGKKALYDIALSWKNKX				
		250	260	270	280	290

Furthermore, ORG75ng-1 shows significant homology to a hypothetical *E.coli* protein:

sp|P45528|YRAL_ECOLI HYPOTHETICAL 31.3 KD PROTEIN IN AGAI-MTR INTERGENIC REGION (F286)
 40 pgil606086 (U18997) ORF_286 [Escherichia coli]
 pgil1789535 (AE000395) hypothetical 31.3 kd protein in agai-mtr intergenic region [Escherichia coli] Length = 286
 Score = 218 bits (550), Expect = 3e-56
 Identities = 128/284 (45%), Positives = 171/284 (60%), Gaps = 4/284 (1%)
 45 Query: 4 KHLQKASDSVVGTLVYVATPIGNLADITLRLALAVLQ KADIICAEDT
 Sbjct: 2 K Q A + S G Y - T P T G I N L A D I T R A L A V L Q D I A E D T R V T A Q L L S A Y G I Q 63
 Sbjct: 2 KQHQSAADNSQ-GQLYTVPTPTGPNLADITQRALEVQAVDIAEADTRHTGQLLQHFGIN 59
 50 Query: 64 GRLVSREHNERQMADKIVGFLSDGLVVAQVSDAGTPAVCDPGAKLARRVREAGFKVPPV 123
 Sbjct: 60 ARLFALHDHNEQQKAETTLAKLQEGQNLAVSDAGTPLINDPGXLVVRTCREAGIRVVPVL 119
 Query: 124 VGASAVMAALSVAGVAESDFYFNGFVPPKSGERRKLPAKWRRAAFFPVMFETPHRIGATL 183
 Sbjct: 120 PGPCAAITALSAAAGLPSPDRFCYEGFLPAKSKGRDALKIAEAEPRTLIFYESTHRLLLDSL 179
 Query: 184 ADMAELFPERR-IMLAREITKTETFLSGTVGEIQTALAADGNQSRGENMVLVLYPAQDK 242
 Sbjct: 180 EDIVAVLGESRYVVVLARELTKTWTETHGAPVGEELLAWVKEDENRRKGEMVLLV-EGHKAQ 238
 60 Query: 243 HEGLSESAQNMILAAELPTKQAAEELAAKITGEGKKALYDIAL 286
 Sbjct: 239 EEDLDPADALRTLALLQAEELPKKAAAALAAEIHGVKKNALYKYAL 282
 65

Based on this analysis, including the presence of a putative transmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 35

- 5 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 295>:

```

1 ATGAAACAGA AAAAACCGC TGCCGCAAGTT ATTGGCTGCAA TGTGGCAGG
51 TTTTGGCGCA GC.AAAGCAC CGGAATTCGA CCGGGCTTTC ..... .
   // .
10 651 ..... . . . GAGTTGG TCAGAAAACC GTTGGAGAG GGTGGAGAC
701 AGGAAAAAAGC CGCGCTTGGAA ATCGATGCC TTTTGGAGA AACCGGTGTC
751 AAACCGTAA

```

This corresponds to the amino acid sequence <SEQ ID 296; ORF76>:

```

1 MKQKKTAAVV IAAMLAGFAA XKAPEIDPAL ..... .
15 201 ..... . . . ELVRNQLEQG LRQEKEARLKI DALLEENGVK
251 P*

```

Further work revealed the complete nucleotide sequence <SEQ ID 297>:

```

1 ATGAAACAGA AAAAACCGC TGCCGCAAGTT ATTGGCTGCAA TGTGGCAGG
51 TTTTGGCGCA GC.CAACAGAC CGGAATTCGA CCGGGCTTTC GGTGGATAACG
101 TGTTGGCGCA GATCATGCAG CAGGGCAGACC GGCAATCGGG GCAGTCCCCA
151 AAACCGGAGC GGCAGCGCAT CGCGAACAGT GCGCTTCGCG GCCTACAAAC
201 TTGGAAAGTT TTGAAAACAC GGGCATTTGGA GGAAAGGTTTG GATAAAGGATA
251 AGGATGTCCA AAACCGCTT AAAATCGCG AAGCGTCTTT TATGGCGAG
301 GAGTTGGCTCC GTTTTCTGG ACGTTGGAA ACGGTTTCGG AAAGACGAGCT
351 GCACAACTGG TTGGAACACGC AAATCCCGAT GATCAAATTC CAGCAGGTCA
401 GCTTCGCAAA CGAAGGAGAC GGGGGTCAGG CGCACGACGT CCTGCTCAAA
451 GGGCTGTCTT TTGAAGGGCT GTATGAAGCTT TATCCGAAAC CGAGCACGGC
501 TTGTTGACGGT TTCAATTATGG CCAGCAGCTT TCCCGAGGCC CTGGCTTCGC
551 AGTTTGGCGC GATGAATCGG GGCAGCTTA CCGGGATCC GTCAAATTG
601 GGCAGAACGGT ATTATCTGTT CAAACTCAGC GAGGTCGGGA AAACACCCGA
651 CGCCGAGCGCT TTGGAGCTGG TCAGAAACCA GTTGGAGCAG GTTTGGAGAC
701 AGGAAAAAAGC CGCGCTTGGAA ATCGATGCC TTTTGGAGA AACCGGTGTC
751 AAACCGTAA

```

This corresponds to the amino acid sequence <SEQ ID 298; ORF76-1>:

```

35 1 MKQKKTAAVV IAAMLAGFAA XKAPEIDPAL VDTLVAQIMQ QADRHAEQSQ
51 KPDQAIRND AVRVLQTEV LKRNALKEGL DKDKDVQNR KIAEASYAE
101 EVRFERLER SEVSDELILKE YEQQIRMLK QQVSFATEER ARQAQQLLLK
151 GLSFEGLMKR YPNDEQAFDG FIMAQOLPEP LASQFAAMNR GDVTDRDVKL
201 GERYTLFKLS EVGKNPFDQP FELVRNQLEQG LRQEKEARLKI DALLEENGVK
40 251 KP*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF76 shows 96.7% identity over a 30aa overlap and 96.8% identity over a 31aa overlap with an ORF (ORF76a) from strain A of *N. meningitidis*:

45		10 20 30
	orf76.pep	/ / / / / / / / / / / /
	orf76a	MKQKKTAAVVIAAMLAGFAAKAPEIDPALVDTLVAQIMQOADRHAEQSQPKDQAIRND
50		10 20 30 40 50 60
		// 70 80 90

-202-

orf76.pep	XELVRNQLEQGLRQEKA RLKIDALI EENGVKPx
orf76a	DVTRDPVKLGERYYLFKLSEVGKNPDAQPFEVLVRNQLEQGLRQEKA RLKIDALI EENGVKPx
	200 210 220 230 240 250

- 5 The complete length ORF76a nucleotide sequence <SEQ ID 299> is:

1 ATGAAACAGA AAAAACCGC TGGCCGAGTT ATTGCTGCAA TTGTTGGCAGG
 51 TTTGGGCC CCGAACAGC CGGAAATCGA CCCGGCTTGTG GTGATACCC
 101 TTGGCCGC CAGATCGAC CAGGGACGCC AGCATTCGGA GCATCTCCAA
 151 AAACCGACG CGGACGCCAT CGGAAACGAT CGGCTCTTCG GCGTCGAAAC
 201 TTGGAACTT TTGAAAACAA GGGCATTTGA GAAAGTTGTT GATAAGGATA
 251 AGGATCTCA AAAACCCGTTT AATAATGCCG AGGCTGCTTT TTGATGCCG
 301 GAGTAGCTTG GTTTCCTGGA ATGGCTGGAA AGGGTTTCGAA AACGGACT
 351 GGCTGACATT TAATGGCCGG AAATTCGGAT GTCAAAATG CAGGAGGTCA
 401 GCTTCGCAAC CGAAGAGGG CCGCCCTCAGG CGCAGCGCT CCTGCTCAA
 451 GGCGTCGTTT TTGAGGGCTT GTAGAGCGCT TATOGCAAGG ACAGGAGCGC
 501 TTGTAACGCT TTCTATTGAG CGCAGCGCT TCCGGAGCGC GTGCTCTTCG
 551 AGTTTCGACG CAGATTAACCG CAAACTCAGC CGGGACCTGG CTTGAAATTCG
 601 GGGCAAGCTT ATTATGCTGTT CAAACTCAGC GAGGTCGGGA AACACCCCG
 651 CGGGACGCTT TTGAGTTG TGCAAGAACCA TTGCAACAAAG GTTGGAGAC
 701 AGGGAAAGG CGCTTGAAATG CTGATGCCA TTGTTGGAAGA AACCGTGTG
 751 AAACCGTAA

This encodes a protein having amino acid sequence <SEQ ID 300>.

1. MKQKPKTAAVY IANLMLGAAKA PKEPDLAPD WDIYQATM QDRHAEQSQ
5. KFDKGQAIRND AVRRLQTLLEV LKNRNLKEGL DDKDKDVQNRD KIAEASATG
10. EYVRFLERSE VTSSESALRQE YERGICRMILKQ QOVSPFESTE ARQACOLLLK
15. GLSFLEGMKYE YPNDEQAFIGF FIMQAQMLPEL LASQFAAMNE GDVTDRPVKL
20. GERYVLFLKS EVGKNPDAQPF FELVRNQLEQ GLQEKEARL IDAILEENGV
25. KP*.

ORF76a and ORF76-1 show 97.6% identity in 252 aa overlap:

- #### 60 Homology with a predicted ORF from *N. gonorrhoeae*

The aligned aa sequences of ORF76 and a predicted ORF (ORF76.ng) from *N. gonorrhoeae* of the N- and C-termini show 96.7% and 100% identity in 30 and 31 overlap, respectively.

-203-

5	orf76.pep	MKQKTTAAVIAAMLAGFAAKAPEIDPAL 	30
	orf76ng	MKQKTTAAVIAAMLAGFAAKAPEIDPALDVTLVAQIMQQADRHAEQSQRPDGQAIRND //	60
	orf76.pep	ELVRNQLEQGLRQEKAIRKLKDALLLEENGVKP	251
	orf76ng	VTRNPVKGLEYYLFKLGAVGKGNPDAQPFELVRNQLEQGLRQEKAIRKLKDALLLEENGVKP	251

The complete length ORF76ng nucleotide sequence <SEQ ID 301> is:

10	1	ATGAACTTCA 51	AAAAGCCG TTGGGCGCA 101	TGGCAGCTT GATCATGC 151	AGACCCGAGC GGCAGGCC 201	TTGGAAGCTT TTGAAACAA 251	AGGATGTC AAACCCGT 301	GAGTAGTCC GGCTGAGTTC 351	TTTGTGCGA GCAACGAC 401	AAATGGCCG AAATCCGCA 451	AAATGGCCG AAATCCGCA 501	AAATGGCCG AAATCCGCA 551	AAATGGCCG AAATCCGCA 601	AAATGGCCG AAATCCGCA 651	AAATGGCCG AAATCCGCA 701	AAATGGCCG AAATCCGCA 751
15	301	ACATTCGGAA GGCGCTTGG 351	GGCGCTTGG GGCGCTTGG 401	GGCGCTTGG GGCGCTTGG 451	GGCGCTTGG GGCGCTTGG 501	GGCGCTTGG GGCGCTTGG 551	GGCGCTTGG GGCGCTTGG 601	GGCGCTTGG GGCGCTTGG 651	GGCGCTTGG GGCGCTTGG 701	GGCGCTTGG GGCGCTTGG 751						
20	351	GGCGCTTGG GGCGCTTGG 401	GGCGCTTGG GGCGCTTGG 451	GGCGCTTGG GGCGCTTGG 501	GGCGCTTGG GGCGCTTGG 551	GGCGCTTGG GGCGCTTGG 601	GGCGCTTGG GGCGCTTGG 651	GGCGCTTGG GGCGCTTGG 701	GGCGCTTGG GGCGCTTGG 751							
25	401	GGCGCTTGG GGCGCTTGG 451	GGCGCTTGG GGCGCTTGG 501	GGCGCTTGG GGCGCTTGG 551	GGCGCTTGG GGCGCTTGG 601	GGCGCTTGG GGCGCTTGG 651	GGCGCTTGG GGCGCTTGG 701	GGCGCTTGG GGCGCTTGG 751								
30	451	GGCGCTTGG GGCGCTTGG 501	GGCGCTTGG GGCGCTTGG 551	GGCGCTTGG GGCGCTTGG 601	GGCGCTTGG GGCGCTTGG 651	GGCGCTTGG GGCGCTTGG 701	GGCGCTTGG GGCGCTTGG 751									
35	501	GGCGCTTGG GGCGCTTGG 551	GGCGCTTGG GGCGCTTGG 601	GGCGCTTGG GGCGCTTGG 651	GGCGCTTGG GGCGCTTGG 701	GGCGCTTGG GGCGCTTGG 751										
40	551	GGCGCTTGG GGCGCTTGG 601	GGCGCTTGG GGCGCTTGG 651	GGCGCTTGG GGCGCTTGG 701	GGCGCTTGG GGCGCTTGG 751											
45	601	GGCGCTTGG GGCGCTTGG 651	GGCGCTTGG GGCGCTTGG 701	GGCGCTTGG GGCGCTTGG 751												
50	651	GGCGCTTGG GGCGCTTGG 701	GGCGCTTGG GGCGCTTGG 751													
55	701	GGCGCTTGG GGCGCTTGG 751														
60	701	GGCGCTTGG GGCGCTTGG 751														
65	701	GGCGCTTGG GGCGCTTGG 751														
70	701	GGCGCTTGG GGCGCTTGG 751														
75	701	GGCGCTTGG GGCGCTTGG 751														

25 This encodes a protein having amino acid sequence <SEQ ID 302>;

1 MKQKQKAAAV AAIAALMGAAV AKAPEAKI PWDVILQVAMIQ QADNRHAQD
 51 RPQGQAIQRND AVRRLQTLVEV LKNRALKEGL DOKDKDQVNQRND KIAEASFYAZA
 101 EYVRFLERSE TWSSESALEP YERQIRNMIKL QOVSFATTEER ARQQALQLL
 151 GLSFLGEMLK YPNDEAQFDG FIMAQOLBEP LASFAGMNE GDVTNRNPVKL
 201 GERYYLEFKLG AVGKNPDAQP FELVRNQLEQ GLRQEKAERLKD IDALLEENGV
 251 KPZ

ORF76ng and ORF76-1 show 96.0% identity in 252 aa overlap

		10	20	30	40	50	60
35	orf76-1.pep	MKQKKTAAVIAAMLAGFAAKAP[EIDPALVDTLVQIMQQADRHA[QSQKPDGQAIRND					
	orf76ng	MKQKKTAAVIAAMLAGFAAKAP[EIDPALVDTLVQIMQQADRHA[QSQKPDGQAIRND	10	20	30	40	50
40	orf76-1.pep	A[VRR]LQTLEV[LKNR]ALKEGLDKDKDVQNRFKIAEAS[YFAE]YVRFLERSETVSEDEL[KP	70	80	90	100	110
	orf76ng	A[VRR]LQTLEV[LKNR]ALKEGLDKDKDVQNRFKIAEAS[YFAE]YVRFLERSETVSE[ALRQF	70	80	90	100	110
45	orf76-1.pep	YEQQIRIMIKLQLQVSFATEEEARQAQQLLKGSLFEGLMKRYPNDEQAFDGFTMAQQLP[P	130	140	150	160	170
	orf76ng	YERQIRIMIKLQLQVSFATEEEARQAQQLLKGSLFEGLMKRYPNDEQAFDGFTMAQQLP[P	130	140	150	160	170
50	orf76-1.pep	LASQFAAGMNRGVDUTRDPVKLGERYYLFKLSEVGKNPDAQPFELVRNQLEQGLRQE[KARLK	190	200	210	220	230
	orf76ng	LASQFAAGMNRGVDUTRNPVKLGERYYLFKLSEVGKNPDAQPFELVRNQLEQGLRQE[KARLK	190	200	210	220	230
55	orf76-1.pep	IDALLEENGVKPX	250				
	orf76ng	IDALLEENGVKPX	250				
60	orf76-1.pep						
	orf76ng						

Furthermore, ORF76ng shows significant homology to a *B. subtilis* export protein precursor.

```

sp|P24327|PRSA_BACSU PROTEIN EXPORT PROTEIN PRSA PRECURSOR >gi|98227|pir||S15269
33kDa lipoprotein [Bacillus subtilis] 33kDa lipoprotein
[Bacillus subtilis]
>gi|2226124|gnl|PID|e325181 (Y14077) 33kDa lipoprotein [Bacillus subtilis]
>gi|2633331|gnl|PID|e1102997 (Z99109) molecular chaperonin [Bacillus subtilis]
Length = 292
Score = 50.4 bits (118), Expect = 1e-05
Identities = 48/199 (24%), Positives = 82/199 (41%), Gaps = 32/199 (16%)
Query: 70 VLKRNRALKEGLDLK-----DKDVQNRFKIAEASF-----YAEEVYRFLERSETVSE 114
      VL      ++ LDK    DK++ N+ K +          Y ++Y++ + E ++++
Sbjct: 53 VLTQLVQEKVLDKKYKVSKIEIDNKLKEYKTQLGDQYTALEKQYKGKDYLKEQVVKYELLTQ 112
Query: 115 SA-----RQFQYERQIRMKILQQVSFATEEARQAQQQLLKGLSPEGILMKRVPN 163
      A      ++++++E   I+ + A ++ A + ++ L KG PE L K Y
Sbjct: 113 KAAKDNKIVTDADIKYEWEGLKGKIRASHILVADKTAEEVEKKLKKGKEFEDLAKEYST 172
Query: 164 DEQAFDQ-----FIMAQQLPEELASLQFAAMMRGDFTRDFVKLGERYYLEFKLSEVGNPDA 218
      D   A   G   F   Q   T   E   +   +   G+V+ DFVK   Y++ K +E   D
Sbjct: 173 DSASASKGGDGLWFAKEGQMDETFSKAFAFKLKTGEVSV-DFVKTQYGWHIIKKTEERGKYDD 231
Query: 219 QPFELVRNQLEQGLRQEKA 237
      EL   LEQ  L   A
Sbjct: 232 MKKELKSEVLEQKLNNDNA 250

```

Based on this analysis, including the presence of a putative leader sequence and a RGD motif in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF76-1 (27.8kDa) was cloned in the pET vector and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 10A shows the results of affinity purification of the His-fusion protein. Purified His-fusion protein was used to immunise mice, whose sera were used for Western blot (Figure 10B), ELISA (positive result), and FACS analysis (Figure 10C). These experiments confirm that ORF76-1 is a surface-exposed protein, and that it is a useful immunogen.

35 Example 36

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 303>:

```

1  ATGAAAAAAAT CTTTCTTAC GCTTGTGTCG TATTCGCTCT TACTTACCGC
51  CAGCGAAATTG GCTTGTACCC TTGGAAATTGG GGATTTGAACCT TTACCCGGCG
101  GCAAAATTTG CGGAAGCTT TGCGCTGACA TTGGTATTGG CTGGCTGTA
151  TCTGTTTGC GGTAAATTAAGG TGACGCGTTTG GTTATGCGG GTTTTTTTG
201  CGTTACGCA TATTGCCAAC AATGTGCAATT ACGGGGATTA TCAAACTGG
251  ATGACG..... .... ..... .... ..... .... ..... .... ..//...
45
1201  ..... CAAACCTAT TCGAGGAGCT GCAAAAGACT CCTGACGGCA
1251  ACTGGCTGTT TGGCTTAACC TCGGATCATG GCCAGTATGT TCGCCAAAGAT
1301  ATCTCACATC AAGGGCACGGT GCAGCCCGAC AGCTATCTCG TGGCGCTGAT
1351  GTTGTACAGC CGGGATATAAGG CGCTGCAACAC GGCCTGGCAAAC CAGGGCTTTTG
1401  CGGCTTGGGA GTTGTGCTTC CATCGACAGC TTCAACAGTT CCTGAATTCAC
1451  ACCTTGGCTT ACGATGATGCC GGTTCAGGT TGTGGCGAAG GCTCGGTAAAC
50
1501  GGGCJACCTG ATTACGGGTG ATGCAGGCG CTTGAGACATT CGCGACGGCA
1551  AGGGGAAATA TGTTTATTCG CAATGA

```

This corresponds to the amino acid sequence <SEQ ID 304; ORF81>:

```

1  MKKSFLTLVL YSSLLTASEI AYPLELGIET LPAAKIAETF ALTFVIAALY

```

51 LFARNKVTRL LIAVFFAFSI IANNVHYADY QSWMT.....
401 ...QTVFEQL QKTPDGWNLF AYTSDDHGQQV RQDINYNQGTQ QPDSSYLWPLV
451 LYSPDKVAQQ AANQAFQAEC IAEHQQLSTF LHITLGYDMP VSGCREGSVT
501 GNLITLGDTQS LNRIRDGKAEY VYPO-

Further work revealed the complete nucleotide sequence <SEQ ID 305>;

1 ATGAAAAAAT CTTCTTCATC GCTTGTCTTG TATTCGCTTAC TACTACCG
 51 CAGCGAAATT GCCTATCGCT TTGATTTGGG GATTAACCCC TTACCCGGCG
 101 CAAAATTCG GAAAGCGTTT GCGCTGACAT TTGATGATTCG TGCGCTGTAT
 151 CTGTTGGC GGTATAAAGGT GACGCCGTTG TTGATGTTGGG TTGTTTTTCG
 201 GTTCAGCATC ATTGGCAACAT ATGGTCATC CGGGCTTATG AAAGAGCTTAC
 251 TGACGGGCAT CAATTTAGTC CIGTAGTGGG AGAACGGTAC GACAGTGGG
 301 AGCGGGGGCCT CGTCGATGTT GGATAAAGTG TTGGCTGCGTC TTGTTGGGG
 351 CGTTGTGGAA CTGATGTTG TTTCGACGCTT TGCGCAATTC CGCCCTAAGTC
 401 CGCCATTTCG TGCCGATATA CTGTTGTCCTT CTTCAATGCTC GTATGATTTC
 451 GTGGCTTCGCT TCAGCACGCA ACAAGGACAGC GGTATTGTTGC CAAACCGAC
 501 ATACAGCCGG ATCAAAAGCCCA ATTATTTCTCG CTTCTGGATTT TTGCTGGAA
 551 GCGGTGTTGCC CTGATCAGTTG TTGATTTAA GCAGGATTTCG CCGCTTAAAG
 601 CAGCTCTTCG CAAGAAAGAT CCGGGACGGC AGTGTCTTAA ATATGCTCTT
 651 GATTAATGGC GAAAGGGCAAG GGGCGGGCGA TTGAGCTGTC TTGTTGCTAC
 701 GACGGAAAC TTGGCGTTT TTAAACCGGC TGCGCAAGCA CGATTATTAAG
 751 CGGATTTGTA ACAAAGATTT TTCCGGCAGGC TTATAGCTG CAGTGFCCT
 801 GCGCAGTTT TTCAATGGGA TACCCGGCAGC CAAACGGCTTG GRACAANCTA
 851 GCGGGCGCGA TACCAAAATAG TTCCGGCTTG CAAAGAGGCA GGCTGTATGAA
 901 AGCTTATTTC ACAGCCGGCA GGGCGGGAAAG GAGATGCGCA TTGTTGACTT
 951 AAATCGTAGA AAAATGGATAG ACCATCTGAT TCAGCGCGAC CAACTTGGCT
 1001 AGCGGCAACGG CGAACATAGCTT CCCGGATGAGA AGCTGTGTC TTGTTGCG
 1051 AAAATCAATC TTGCGAGGAGG CRAGCATTTT ATCGTTGTC AACAAACCGG
 1101 TTGCGAGGCC CTACAGCGCA CATTGTTGCA GCTTCAAGAT AGGATTATGG
 1151 GGAAAGCGGA TATTGTTGGAA ATAACGACAA CACCATCCTA CAAAGAGGAA
 1201 CAAATGATTC AACACCTATG CGACGGACGTC CAAAGACGCG TGCGGCGCA
 1251 CTGGCTGTITT GCCTATACCTT CGCGATCTGG CAGTATGTC CGCCAGATA
 1301 TCTACATACTT AACGGACGCTG CGGCCGACAA GCTATCTGGT CGGCCCTAGT
 1351 TTGTCAGACCC CGGATTAAGGG CGTCGACAGC GCTGGCAACAG AGGCTTCTG
 1401 GCTCTGGCGG ATTCGCTTCATC ATCAGCAGCTT TTCAAGCTTC TTGATCACA
 1451 CGTTGGCGCA CGATAGCGCA GTTTCAGCGTTT CGCCGGAGGA CGTGGGAAACG
 1501 GGCAAGCTGAA TTACGGGTTG TGCGAGGACG TTGAAACATTG CGCAGCGGCA
 1551 GGCGACGTTG GTTTCAGCGC ATGTA

This corresponds to the amino acid sequence <SEQ ID 306: QRE81-1>:

40	1 MKKSFLTQLVL YSSLTASEI AYRFVFGIET LPAAKIAETF ALTFVIAALY 51 LFARYKVTLR LIAVFFAFAI IANNVHVAY QCSWNTGINYW MLNTEKEVGT 101 SASAGMLDRL WLPVLWGLVIR KPLRFTCSLAFK RRRTFHSDLFZ LAFIMMIF 151 VRSFDITQKGS GIKPSKTYFV IKANFYSFGY FGVRLVYQQL DRFLSRAPIKF 201 QPAPSKIGQG SVQNIVLIMR ESESAHHLK FGYYGETSPF ULRSLQADKF 251 PIKVQSISAG FMTAVLSVPLF NFIAPNHGL EQISGQDDTNF RFLKEAGYEF 301 TYYTFSQAQO EMALINLGE KWIDHLQFT QLGYNGNDNF PDEKLHLPLD 351 KINLQQQKHFJ IVLHQRGQSH FAJGLPQDOP KVFGAEADYV KINTD1HKTD 401 QM1QTVFQEQL QKQFDGWNLFV ATTSHDHQVY RDILQYNTQGP QDSYLVPLV 451 LYSPLKAVQO AANQAFAPCE JAFHQLQSTF LIHTLGYMPP VSQGREGSVT 501 GNLITGDAQS LNRDQKAEY VYPO-
----	--

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF81 shows 84.7% identity over a 85aa overlap and 99.2% identity over a 121aa overlap with an ORF (ORF81a) from strain A of *N. meningitidis*:

-206-

	orf81.pep	LIAVFFAFSIIANNVHYADYQSWMT ::: : :				
5	orf81a	LIAVFFAFSIIANNVHYAVQSWITGINYWLMILKEITEVGAGASMLDKLWLPAIWGVLE 70 80 90 100 110 120	//			
	orf81.pep			120	130	140
	orf81a			QTVFEEQLKTPDGWLFAYTSDHGQYVRQD : : : : : : :		
10	orf81a	IPHANGLEQISGGDIDVVKYDNTIHKTDQMIGTWFEEQLQKOPDGWLFAYTSDHGQYVRQD 280 290 300 310 320 330				
	orf81.pep		150 160 170 180 190 200			
15	orf81a	IYNQGTVPQDSYLVLPLVLYSPDKAVQQAANQAFAPCEIAFHQLSTFLIHTLGYDMPVSG 340 350 360 370 380 390				
	orf81.pep		210 220 230			
20	orf81a	CREGSVTGNLITGDAGSLNIRUDKASYVYPOX 400 410 420				

The complete length ORF81a nucleotide sequence <SEQ ID 307> is:

25	1	ATGAAAAAAAT	CCTTTTCGTTCTCTTCTG	TATTCGTC	TACTTACTGC	
	51	CAGCGAATTG	GCTTATTCGCT	TTGTTATTCG	AATTGAAACCC	TTACCGGCTG
	101	CAAAATGGC	AGAACGCTT	GGCGTGCACAT	TTGTGATTCG	TGCCGCTGAT
	151	CTGTTTGGCC	GTTATAAGGC	AACCGCTTG	TTGATTGCGG	TTGTTTTCCG
30	201	GTTCACCAT	ATTGCCAACAA	ATGTCATTA	CCGGGTTTAT	CAAAGCTGGA
	251	TAACCGGG	TAATTATTTGG	CTGGATGCTTA	AAAGGATTAC	CAAAGTTGGC
	301	GCCCCAGGG	CTGGCTATGTT	GGATAATGTT	TGGCTGCCTG	CTTGTGTTGGG
	351	CGTGTGGAA	GTCAATGTTG	TTTCAGCCTG	TGCCAAGTTC	CCCCGTAAGA
	401	CGCATTTTC	TGGCCATATA	CTGTTGGCT	TCTTAATGCT	GATGATTTC
	451	GTGGCTCC	TCGACACCAA	ACAAAGAACAC	GGTATTTGCG	CCAAACCGAC
	501	ATACAGCCG	ATCAAGGCCA	ATTAATTCG	CTTCGGTTAT	TTTGTGCGAC
	551	CGCTGTTGG	TTATGATGTC	TTGATTTTA	GCAGAGATCC	TGTGTTCAAA
	601	CAGCGCTGTC	CAAGCGAAGT	CGGGCAAGGC	AGTATTTCAA	ATATCGCTCT
	651	GATTATGGC	GAAGCGCAA	GGGGCGGCCA	TTTGAATATG	TTTGGCTACG
	701	GGCGGAAAN	TTGGCGCTT	TTGAGCCACG	TTTCCGCAAGG	CGATTTTAAAG
35	751	CGGATTGTA	AACAAAGCTA	TTCCGCAGGC	TTTATGACGG	CACTATCCT
	801	GGCCAGCTT	TTTAACTGCTA	TAACCGCATTC	CAACGGCGTC	GAACAAATCA
	851	GGCGGCGCG	TATTGTTGGAT	AAAGTACGCCA	ACACCATCCA	CAAAACCGAC
	901	CAAAATGATT	AAACCGTATTG	CGAACGACTG	CAAAGCAGC	CTGACCGGCCA
	951	CTGGCTGTT	GGCTATACCT	CGGATCATGG	CCAGTAGTCG	CCCGCAAGATA
40	1001	TCTCAATCA	AGGCCAGCTG	CAGCCCGACAA	GCTATCTCGT	GGCCGCTGTTG
	1051	TTGTCAGCCG	CGGATACGGC	CTGCAACAGG	GGGCTTTTGC	
	1101	GCCTTGGCA	ATTCAGCTTC	ATCAGCAGCT	TTCAACGTTG	CTGATTCAACA
	1151	CGTTCGGCTA	CGATATGGCG	TTTCAGGTT	GTGCGCAAGG	CTCGGTAACCG
	1201	GGCAACCTCA	TTACGGCTGA	TCGACGGACG	TTGAACATTC	GGCACGGCCA
	1251	GGCGAAATA	GTTCATCCG	AATGA		

50 This encodes a protein having amino acid sequence <SEQ ID 308>:

1	MKKSLFVFL	YSSLLTASEI	AYRFVFGIET	LPAAKMAETF	ALTFVIAALY	
51	LFARYKATRL	LIAVFFAFSII	ANNVHYAVY	QSWITGINYW	IMLKIEITVEG	
101	GAGASMLDKL	WLPALWGVL	EMLFCSLAKF	RRKTHFSADI	LEFAFLMLMIF	
151	VRSFDTKRKH	GISPKPTYS	IKANYFSFGY	FVGRVLPYOL	FDLSKIPVFK	
201	QPAPSRRIGQC	SQINVLIMG	ESESAHHKL	FGYGRETSPF	LTLQSLQADF	
251	PLVQSYSSAC	FMTAVSLPST	FNVPIHANGL	EQISGGDIVD	VKDNTIHKID	
301	QMIQTVFEEQL	QKOPDGNNW	AYTSDHQCVV	RQDIYNQGTV	QPDSYLVPLV	
351	LYSPDFKAVQQ	AANQAFAPCE	IAFHQLSTF	LIHTLGYDMP	VSGCREGSVT	
	401	GNLTIDGADS	LNIRDGKAEY	YIFP*		

60 ORF81a and ORF81-1 show 77.9% identity in 524 aa overlap:

60	orf81a.pep	10	20	30	40	50	60
	orf81a	MKKSLFVFLYSSLLTASEIAYRFVFGIETLPAAKMAETF	LPAAKMAETF	ALTFVIAALY	LFARYKATRL		
65	orf81-1	MKKSLFTLIVLYSSLLTASEIAYRFVFGIETLPAAKIAETF	ALTFVIAALY	LFARYKVT	RL		

		70	80	90	100	110	120
	orf81a.pep	LIAVFFAFSIIANNVHYAVYQSWITGINYWLMLKEITEVGAGASMLDKLWLPALMGVLE					
5	orf81-1	LIAVFFAFSIIANNVHYAVYQSWMTGINYWLMLKEVTEVGSAGASMLDKLWLPALMGVLE	70	80	90	100	110
	orf81a.pep	130	140	150	160	170	180
10	orf81-1	VMLFCSLAKFRRKTHIFSDADILFALMLMIVFVRSFDTKQEHEGISPCKPTYSRIKANYFSFGY	70	80	90	100	110
	orf81a.pep	VMLFCSLAKFRRKTHIFSDADILFALMLMIVFVRSFDTKQEHEGISPCKPTYSRIKANYFSFGY	130	140	150	160	170
15	orf81-1	130	140	150	160	170	180
	orf81a.pep	FVGRVLPLYQLFDLSKIPVFKQAPRIGQGSIQNIVLIMGESESAAHLKLFGYGRETSPP	190	200	210	220	230
	orf81-1	FVGRVLPLYQLFDLSRIPAFKQPAPKSKIGQQGSVNQIVLIMGESESAAHLKLFGYGRETSPP	190	200	210	220	230
20	orf81a.pep	190	200	210	220	230	240
	orf81-1	LTQLSQADFKPIVKQSYSSAGEMTAVSLSPSFVNIPHANGLEQISGGD-----	250	260	270	280	
25	orf81-1	LTRLSQADFKPIVKQSYSSAGEMTAVSLSPSFVNIPHANGLEQISGGDTNMFRLAKEQGYE	250	260	270	280	290
	orf81a.pep	250	260	270	280	290	300
	orf81-1	-----					
30	orf81a.pep	TYFYSAQAENEMAIIINLIGKWKIDHLLQPTQLGYGNQDNMPDEKLLPLFDKINLQQGKHF	310	320	330	340	350
	orf81-1	310	320	330	340	350	360
	orf81a.pep	290	300	310	320		
35	orf81-1	IVLVHQRGSHAPYGALLQPQDKVVFGEADIVVKYDNTIHKTQDMQIQTVEFOLQKQPDGNWL	370	380	390	400	410
	orf81a.pep	290	300	310	320		
40	orf81-1	370	380	390	400	410	420
	orf81a.pep	330	340	350	360	370	380
	orf81-1	AYTSDHGQYVRQDIYNQGTVPQPSYLVPLVLYSPDKAVQQAANQAFAPCEIAFHQQLSTF	430	440	450	460	470
45	orf81a.pep	330	340	350	360	370	380
	orf81-1	AYTSDHGQYVRQDIYNQGTVPQPSYLVPLVLYSPDKAVQQAANQAFAPCEIAFHQQLSTF	430	440	450	460	470
	orf81a.pep	390	400	410	420		
	orf81-1	390	400	410	420		
50	orf81a.pep	LIHTLGYMPVSGCREGSVTGNLITGDAGSLNIRDGKAEVYVYPOX	490	500	510	520	
	orf81-1	LIHTLGYMPVSGCREGSVTGNLITGDAGSLNIRDGKAEVYVYPOX	490	500	510	520	

50 Homology with a predicted ORF from *N.gonorrhoeae*

The aligned aa sequences of ORF81 and a predicted ORF (ORF81.ng) from *N. gonorrhoeae* of the N- and C-termini show 82.4 % and 97.5% identity in 85 and 121 overlap, respectively:

	orf81.pep	MKKSFRTLVLVYSSLTASEIAYPLEGLIETLPAAKIAETFALTIVIAALYL FARNKVTRL	60
55	orf81ng	MKKSLFVLFLYSSLTASEIAYRFVFGIETLPAARMAETFALTIVIAALYL FARYKASRL	60
	orf81.pep	LIAVFFAFSIIANNVHYADYQSWMT	85
60	orf81ng	LIAVFFAFSMIANNVHYAVYQSWMTGINYWLMLKEVTEVGSAGASMLDKLWLPALMGVRE	120
	orf81.pep	//	
	orf81ng	QTVFQEQLQKTPDGNWLFAYTSDHGQYVRQD	433
	orf81ng	ALLQPQDKVVFGEADIVVKYDNTIHKTQDMQIQTVEFOLQKQPDGNWL	433
65	orf81.pep	FAYTSDHGQYVRQDIYNQGTVPQPSYLVPLVLYSPDKAVQQAANQAFAPCEIAFHQQLSTF	493
	orf81ng	IYNQGTVPQPSYLVPLVLYSPDKAVQQAANQAFAPCEIAFHQQLSTF	493

orf81.pep	CREGSVTGNLITGDAAGSLNIRDGKAAYVVPQ	524
	::: : : : : :	
orf81ng	CREGSVTGNLITGDAAGSLNIRNGKAAYVVPQ	524

The complete length ORF81ng nucleotide sequence <SEQ ID 309> is:

5	1 ATGAAAAATC CCCTTTTCGT TCTCTTTCGT TATTCCATCCC TACTTACCGC
	51 CAGGGAAATC GCTCATTCGT TTGTTATTCGG AATTGAAACCT TTACCCGGCTG
	101 CAAAAATGGC GGAAAGCTTG GGGCTGACAT TTATGATTTG TGCGCTGTAT
	151 CTGTTTTCGGC GTTATAAAGGC GGGCGGCGCTG CTGATTTGGCG TTGTTTTTCGC
	201 GTTCAAGCATG ATTGCCAACAA ATSGCATTAA CGCGGTTTAT CAAAGCTCGA
	251 TGACGGGTAT TAACATTG TGATGCTGA AAGAGGTTTAC CGAAGTCGGC
	301 AGGGGGGGCG CGTGGTGTCTT GGTAATCTTG TGGCTGCGTGC CTTTGTCGGG
	351 CCTGGCGGAA GTCATCTTGT TTTCGAGCTT TGCCAAGTTG CGCCCTAAAGA
	401 CGCATTTTC TGCGATATA CTGTTGCGCT TCCATAATGCT GATGATTTC
	451 GTGGGCTCGT TCAGCACAGCA AAAGAGCAC GGTTATTGCG CCAAACCGAC
	501 ATACAGCGC ATCAAAGCCA ATTATTTCTAC CTTCGGTTTAT TTGTCGGGC
	551 GGCTGGTGGC GTTACATGGT TTTCGATTTAA GCAAGATCC TTGTTGTCAAA
	601 CAGCCCTGCTC CAAGAACAAAT CGGGCAAGGCG AGTATTCAAA ATATGTCCT
	651 GATTATGGGC GAAAGGGGAA CGGCGCGCA TTGAAATTG TTGGTTACG
	701 GGGCGGAAAC TTGGCGTTT TTAAACCGGC TGTCGCAAGG CGATTTTAAAG
	751 CGCGATTTGCA AACAAAGTTG TTCCCGACGGC TTATGAGCGC CGTATCCCT
	801 GCGCAATTTC TTAACTGTC TAACCGACCG CAAACGGCTT GAACAAATCA
	851 GGGGGGGCGA TACCAATGTT TTCCCGCTCG CAAAGAGCGC GGCGTATGAA
	901 ACGTATTTT ACAGCTGCCA GGCCTGAAAAC CAAATGGCAA TTGTTGAACTT
	951 AATGGGTAAG AATGGATAG ACCATCTGAT TCAGCCGAGC CAACTTGGCT
	1001 ACGGGGAAAC CGAACATATG CCCGATGAGA AGCTGCTGCC GTTGTTCGGAC
	1051 AAAATCAATTG TGACGAGGG CAGGCATTIT ATGCTGTTGC ACCAAACGGG
	1101 TTGCAAGGCC CCATACGGGC CATTTGCGA GCCTCAAGAGT AAAGTATTGCG
	1151 GCGAAAGCGCA TATTGTTGGT AAAGTACGACA ACACATCCA CAAACACGGAC
	1201 CAAATGATTC AACACCTTGT CGAGGAGCTG CAAAAGCAGC CTGACGGCAA
	1251 CTGGCTTTT GCTTACATCT CGATGATCGG CGTATATGTT GCCTCTGGTT
	1301 TCTACAACTCA AGGCACGGTG CAGCCCCACA GCTATATGTT GCCTCTGGTT
	1351 TTGTACAGCCG CGGATTAAGGC CGTGCARACG GTGCGCAACC AGGCTTTTGC
	1401 GCTTGGGGAG ATTGCTCTTC ATCACAGCTT TTCAACGTTG CTGATTACAA
	1451 CGTTGGCTCA CGATAGCGG GTTTCAGGTT GTGCGCAAGG CTCGGTAACA
	1501 GCGAACCTGA TTACGGGGCA TGACGGAGC TTGAAACATTG GCAACGGCAA
	1551 GCGGGAAATG GTTTATCCCG AATAA

This encodes a protein having amino acid sequence <SEQ ID 310>:

1	MKKSLFVLFL YSSLILTASEIAYRFVFGIETLPAAKMAETFALT	ALTMIAALY
	LEFARYKASRL LIAVFFAFSE MIAINNHAVY OSWMTGINYW LMILKEVTEVG	
40	101 SAGASMLDKL WLPLGVVAE VMLFCSLAKRHE KRRTHFSADI LFAPFLMLMIF	
	151 VRSFTVTKOEH GISPKPTVSYI IKANYFSSGY FVRGVLPVQL FDLFSKIPVFK	
	201 QPAPSKIGQ SQTINIVLIMG ESESAAHLK1 FGYGRETSPF LTRLSQADFEK	
	251 PIVKQSYSG FMFTAVSLPSF FNVI:PHANGL EQISGGDTNM FRILAKEQGYE	
	301 TYFTSYAQSEN QMATAILNLIGK KWDHLLIQPT QLYGNGDNM PEDEKLLFLFD	
	351 KINLQGQRHF IVLHQRGSHS PYGALLQPDQ KVFGAEADIVW KYDNTNIHKTID	
	401 QM1QTVQHDFV IVLHQRGSHS PYGALLQPDQ KVFGAEADIVW KYDNTNIHKTID	
	451 LYSPIRKAVOO AAHQAFACE IAFHQQLSTF LIHTLGYDMP VSGRCGSEVT	
	501 GNLLITGAGS LNAIRNGKAEY VVPQ*	

ORF81ng and ORF81-1 show 96.4% identity in 524 aa overlap:

50	orf81ng-1.pep	10 20 30 40 50 60	
	MKKSLFVLFLYSSLILTASEIAYRFVFGIETLPAAKMAETFALT	FMAALYL	
	LEFARYKASRL LIAVFFAFSE MIAINNHAVY OSWMTGINYW LMILKEVTEVG	FARYKASRL	
	orf81-1	10 20 30 40 50 60	
55	orf81ng-1.pep	70 80 90 100 110 120	
	LIAVFFAFSE MIAINNHAVY OSWMTGINYW LMILKEVTEVG	SAGASMLD	
	orf81-1	70 80 90 100 110 120	
60	orf81ng-1.pep	130 140 150 160 170 180	
	VMLFCSLAKRHKRTHFSADI LFAPFLMLM FVRSFDTKQEHGISPKPTYSRIKANYFSSGY	WLPLGVVAE	
	orf81-1	130 140 150 160 170 180	
65	orf81ng-1.pep	VMLFCSLAKRHKRTHFSADI LFAPFLMLM FVRSFDTKQEHGISPKPTYSRIKANYFSSGY	WLPLGVVAE
	orf81-1	VMLFCSLAKRHKRTHFSADI LFAPFLMLM FVRSFDTKQEHGISPKPTYSRIKANYFSSGY	WLPLGVVAE

		130	140	150	160	170	180
5	orf81ng-1.pep	190	200	210	220	230	240
	orf81-1	FVGRVLVPYQLFDLSKIPVFQKQAPSKIGQQSIQNIVLIMGESESAAHLKLFGYGRETSPPF					
		190	200	210	220	230	240
10	orf81ng-1.pep	250	260	270	280	290	300
	orf81-1	LTRLSQADFPIVKQSYSAGFMTAVSLPSFVNVIIPHANGLEQISGGDTNMFRLAKEQGYE					
		250	260	270	280	290	300
15	orf81ng-1.pep	310	320	330	340	350	360
	orf81-1	TYFVSAQAEENQMAILNLIGKKWIDHLLQPTQLGYGNNDNPDEKLLPLFDKINLQQGRHF					
		310	320	330	340	350	360
20	orf81ng-1.pep	370	380	390	400	410	420
	orf81-1	IVLHQRGSHAPYGAALLQPDKVKGFEADIVDKYDNTIHKTDQMIQTVEFQLQKQPDGNWL					
25	orf81-1	IVLHQRGSHAPYGAALLQPDKVKGFEADIVDKYDNTIHKTDQMIQTVEFQLQKQPDGNWL					
	orf81ng-1.pep	370	380	390	400	410	420
30	orf81-1	430	440	450	460	470	480
	orf81-1	AYTSDHGQYVRQDIYNGQTVPQDSYLVLVLYSPDKAVQQAANQAFAPCEIAFHQLSTF					
		430	440	450	460	470	480
35	orf81ng-1.pep	490	500	510	520		
	orf81-1	LIHTLGYMPVSGCREGSVTFGNLITGDAGLNIRNGKAEVYVPOX					
		490	500	510	520		

Furthermore, ORF81ng shows significant homology to an *E.coli* OMP:

40	gi 1256380 (U50906) outer membrane adherence protein-associated protein [E. coli] Length = 547
	Score = 87.4 bits (213), Expect = 2e-16
	Identities = 122/468 (26%), Positives = 198/468 (42%), Gaps = 70/468 (14%)
45	Query: 25 VFGIETLPAAKMATAFFA-LTMFIMIALYLKFYRKAAS--RLLIAVFFAFSMIANNVHYAQY 81
	Sbjct: 29 VFGITNLVASSGAHMVQRLLFFVLTILVVKRISSSLPLRLLVAAPFVL-LTAADMSISLY- 86
50	Query: 62 SWMT-----CINYWLMILKEVTEVGSGAGSMMLDKLNLPALWGVAEVMLFCSLAKFRRKT 134
	SW T + + + + EV A ML ++ P L A + L +
	Sbjct: 67 SWCFTGFTFNGDAFISVLQSDPDEV---ARMILG-MYSPLCAFAPLSSLFLAVIYKV 141
55	Query: 135 HWSADILFALMILMFVRSF-----DTKQEHGSPKPTYSRIKAN--YFSPFGYVG 183
	+ L+L+ S K D ++ SP SR +F+ YF
	Sbjct: 142 SLPTKKVTLGILVVISGSLPSACQFAYDKANKNAFSPYLASRPFATYTPFNLYFV 201
60	Query: 184 RVLVYQ---LFDSLKIFVFKQKAPSKIGQQSIQNIVLIMGESESAAHLKLFGYGRETSPL 241
	+Q L + + +F+ + I VLI+GES ++ L+GY R T+P +
	Sbjct: 202 AAKEHQRLLSIANTVYFQL---SVRDTGIDTVVLIVGESVRVDNMSLYGYTRSTTFQV 257
65	Query: 242 TRLSQADPFPIVKQSYSAGFMTAVSLP---SFFNVIIPHANGLEQISGGDTNMFRLAKEG 298
	+O + + Q + S TA+S+ + + +V H I N+ +A + G
	Sbjct: 258 E--AQRQIKLNFNAIQSGAPYTAZLSPVSLTADSVLSH----DIHNPYDNIINMANQAG 310
70	Query: 299 YETSYKSAQAA---ENQMAILNLIGKKWIDHLLQPTQLGYGNNDNPDEKLLPLFDKINL 355
	+T++ S+Q+ +N A+ ++ + + + V G DE LLP + Q
	Sbjct: 311 FOTWLSSQVSAFRONGTAVTSI-----AMRAMETVYVRFG---DELLPHLSQLQQ 359
	Query: 356 --GRHRFVHLQRGSHPAYGALLQPDKVGFGEADIVDK-YDNTIHKTDQMIQTVEFQLQ 412
	+ I VHLW GSH + + VF D D YDN+H TD ++ VFE L+
	Sbjct: 360 NTQQQKQLLIVLHLNGSHEPACSAQYQSSAFAVQFQDDQDACYDONSIIHYYTDSLLGQVFELLK- 418

Query: 413 QPDGNWLFAYTSDHGG--QYVRQDIYNQG--TVQPDTSYIVPL-VLYSP 454
 D Y +DHG +++Y G +Y VP+ + YSP
 Sbjct: 419 --DRRASVMYFADHGLERDPTKRNVYFHGGREASQQAYHVPMFIWYSP 464

5

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

10 Example 37

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 311>:

1 ...ACCCGTCTCC TCTTCATCCC CCTCGTCTTC ACAC GTGCC GCACACTGAC
 51 CGGCATPACT GCCCAAGCCG GCGGAAACCG CTTTGCGCTC GAACAGAACG
 101 TGCTGGCGCG ATCTGCGCG GCGGCGCTCA AAGAAATGGA TTTCGCCCC
 151 YTAAGAACGAC GCAAGACCGG CYTTTACCTG TCCGTTAATGG CGGACCAAGG
 201 CGCGGTACAC TATAAGGGGG GACGCTACTC TATCGAGCGA CTGATACCCG
 251 TACCGTACCA CAACAAACCC GAAAGTGCCA CCAAAATACAG CTACCCCCC
 301 TACCGACTA CCTGGCACCA CAAATCCGAC CGCGTCTCCA CGCTTACCCAC
 351 TTCCACATTCG CTTTGAACG CCTCCCGCGC CGCVCYTGACG AAAAACACCG
 401 GACGCCAAGGG CGAACCGCTC CGCCGACTGT CGCTCAACGG CACGGGGGAC
 451 TACCCGACCG AAAGGACTCTG CGCCGACCCCG CGCGACGGTT CCTTCTCTGAC
 501 CAACCTCATC CAAACCGCTC TCTAACCTCG CGCGATCAGA GTCgtTACCGC
 551 CGCrATACCGC CGACACCCGAC GTATTCTCAA CGCTCGACCT A...

This corresponds to the amino acid sequence <SEQ ID 312; ORF83>:

25 1 ...TLLFIPLVL TXCGTLLTGTIL AHGGGKRFAV EQLELVAASSR AAVKEMDLSA
 51 LKGCCRQVYY SVMGDQSGSN ISGGRYSIDA LTRGGYINNNN ESATQSYSPA
 101 YDTTATTKSD ALSSVTSTLS LNAPAXLIT KNSGRKERS AGLSVNGTGD
 151 YRNETLLAND RDVSFLTNLI QTVFYLRGIE VVPPXYADTD VFVITVDV..

Further work revealed the complete nucleotide sequence <SEQ ID 313>:

30 1 ATGAAAGAACCC TGTCTCTCTC CATCCCTCTC GTCTCTCACAG CCTGGGGCAC
 51 ACTGCGCGC ATACCGCCG CCTGGCGCGG CAAACGTTT GCGCTGAGAC
 101 AAGGACTCTG CGCCGCGATCG TCCCGCGCGC CGCTGCAAGA ATGGGATTTG
 151 TCCGGCCCTAA AGGGGCGCAAA AGGCCGCGCTT TACGCTCTCG TTATGGGGA
 201 CCAAGGTTCCG GCGAACATAA CGCGCGGAGG CTACTCTCTAC GAGGCACTGA
 251 TAGCGCGGG CTACCAACCA AACCCGAAA GTGCCACCAAA ATACAGCTAC
 301 CCCGGCTACG AGACTACCCG CAAACCGAA TCGCGAGCGC TCTCCAGGCT
 351 AACACTTCCC ACATGCGTTT TGAACGCCCC CGCGCGCGC CTGAGAAAAA
 401 ACAGCGGAGC CAAAGCGGA CGCTCCGGC GACTGTCTGT CAAAGCGCAC
 451 GGGGACTTACCG CAAAGCGGAAC CGCTCTCGGC AACCCCGGGG ACCTTTCTT
 501 CCTGGACAC CTCATCCAA CCTCTCTCTA CCTGGGGGGC ATCGAAGTCC
 551 TACCCGGCGG ACAGCGCGAC CGCGCTGAT CGTAGAACCTT CGAGCTATTC
 601 GGACCGCTCG CGACCGCTCA CGACTGTGAC CTCTACAAAGG CGGAAACCT
 651 TAACAGCCCAA ACCGAAGCTGG ATATTTGCG CGTTGACCCG GACAGCCGG
 701 AACCTCTGAT TACCCCTAA ACCGGCGCTT CGGAATCTCA ATACCAAGAA
 751 CAAATGCGCC TTGGGACCGG CCTTACAAAT GTCAAGAAAAA CGCTCAAAGC
 801 CTCAGACCGCG CTGATGCTCG ATTCTCTCGA CATTACCCCCC TACGGGGACA
 851 CAACCCCGCA AAACCGTCCC GACTTCAACAC AAAACAAACGG TAAAAAACCC
 901 GATGTGCGCA ACGAAGTCAT CGCCGCGCGC AAAGGAGGAT AA

This corresponds to the amino acid sequence <SEQ ID 314; ORF83-1>:

50 1 MTLILLIPL VLTACGTLLTGP IPAHGGGRF AVEQELVAAS SRAAVKEMDL
 51 SALRKRAAL YVSMGDQGS NSISGGRYSI DALIRGGYINN NPESATQSY
 101 PAYDTTATTKSD ALSSVTSTLS TSLLNAFAAAA LTKNNSGRKRE RSAGLSVNGT
 151 GDYRNETLLAND RDVSFLTNLI LIQTFYLRGIE IEVVVPPEYAD TDVFVITVDV
 201 GTVRSRTELH LYNAETLKAQ TKLEYFAVDR DSRKLITPK TAAYESQYQE
 251 QYALWTGPyY VSKTKVAKASDR LMVDFSIDTP YGDTTAQNRP DFTQNNGKPK

301 DVGNNEVIRR R KGG*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF83 shows 96.4% identity over a 197aa overlap with an ORF (ORF83a) from strain A of *N.*

5 *meningitidis*:

	10	20	30	40	50
orf83.pep	TLLLFPLVLTXCGTLTGILAHGGGKRFAVEQELVAASSRAAVKEMDLSALKGRKAAX				
orf83a	MKTLLXLIPPLVLTAACGTLTGIPAHGGGKRFAVEQELVAASSRAAVKEMDLSALKGRKAAL	10	20	30	40
10					50
orf83.pep	YVSVMGDQSGNISGGGRYSIDALIRGGYHNPNESATQYSYPAYDTTATTKSDALSSVTTS	60	70	80	90
orf83a	YVSVMGDQSGNISGGGRYSIDALIRGGYHNPNESATQYSYPAYDTTATTKSDALSSVTTS	70	80	90	100
15					110
orf83.pep	TSLLNAPAXLTKNSGRKGERASGLSVNGTGYDRNETLLANPRDVSFLTNLIQTVFYLRG	120	130	140	150
orf83a	TSLLNAPAAALTKNSGRKGERASGLSVNGTGYDRNETLLANPRDVSFLTNLIQTVFYLRG	130	140	150	160
20					170
orf83.pep	IEVVPPPEYADTDVFVTVDFVFTVRSRTELHLYNAETLKAQTKLEYFAVDRDSRKLLIAPK	180	190		
orf83a	IEVVPPPEYADTDVFVTVDFVFTVRSRTELHLYNAETLKAQTKLEYFAVDRDSRKLLIAPK	190	200	210	220
25					230
					240

The complete length ORF83a nucleotide sequence <SEQ ID 315> is:

30	1 ATGAAAACC TGCTCTNCCT CATCCCCCTC GTCCCTCACAG CCTGCCGCAC
	51 ACTGACCGGG ATACCCCGCC ACAGCGCGG CAAACGCTTT CGCGCTGAC
	101 AAGAACCTGT CGCCGCATCG TCCCGGCGC CGCTCAAAGA AATGGACTTG
	151 TCCCGCCTGA AAGACGCAA AGCCGGCCCT TAGCTCTCG TTATGGGCA
	201 CCAAGCTTGC GGCAACATCAA GGGCGGAGC CTACTCTATC GACCCCACTGA
	251 TACCGGGCGG CTACACCAAAC AACCCGGAAA GTGCCACCAAA ATACACCTAC
	301 CCCGCTTACAG ACACATACCG CACCAACCA TCCGACGGCGC TCTCCACAGCT
	351 AACACTTCG ACATCGCTT TGAACGGCCC CGCCCGGCC CGTACGAAAAA
	401 ACACGGGGAG CAACAGGCAA GCGTCGCGC GACTGTCCGT CAACGGCAGC
	451 GCGGCTTACCC GCAACGAAAC CCTGCTCGGG AACCCCCCGG AGCTTCCCT
	501 CCTGACCAAC CTACATCCAA CGCTCTCTCA CCTCGGGGGG ATCGAACAGTG
	551 TACCGGGCGA ATACCGCGAC ACACGGGAT ACCTGACCGT CGACGAGCTTC
	601 GGCACCGTGC GCACGGCGAC CGAACCTGCAC CTCTACACGG CGAGAACCT
	651 TAAAGCCCAA ACCAACGTCG AATATTCGGC CGTGTACCGT GACAGCGGA
	701 AACCTGCTGAT TGCCCCCTAAAC CGACGGCGCT ACAGAACCCA ATACCAAGAA
	751 CAATAGCGGC TGCGCTTAAAC CGACGGCGCT ACAGAACCCA ATACCAAGAA
	801 CTGAGACCGC CTGAGTGGCCTG ATTATCTCCGA CATCACCCCC TACCGGGACA
	851 CAACCGCGCA ACCACGTCG CACTGCAAC AAAAACACGGG TAAATACCGC
45	901 GATGTCGGCA AAGAACGTCAT CGGCCCGCG AACGGAGGAT AA

This encodes a protein having amino acid sequence <SEQ ID 316>:

50	1 MKTLLXLIPPL VLTAACGTLTG IPAHGGGKRFAVEQELVAAS SRAAVKEMDL
	51 SALGRKAAL YVSVMGDQGS GNISGGGRYSI DALIRGGYHN NPESATOYSY
	101 PAYDTTATTK SDALSSVTTS TSILLNAPAAA LTKNNSGRKGE RSAGLSVNGT
	151 GDYRNELLAA NRPDVSLFTN LIQTVFYLRG IEVVPPPEYAD TDVFVTVDFV
	201 GTVSRTELTH LYNAETLKAQ TKLEYFAYDVR DSRKLIIAPK TAAYE8QYQE
55	251 QYALWMGPIS VGGTKVAKSDR LMWVFSDDITP YGDTTAQNRF DFKQNNNGKKP
	301 DVGNNEVIRR R KGG*

ORF83a and ORF83-1 show 98.4% identity in 313 aa overlap:

	10	20	30	40	50	60
orf83a.pep	MKTLLXLIPPLVLTAACGTLTGIPAHGGGKRFAVEQELVAASSRAAVKEMDLSALKGRKAAL					

	orf83-1	PLVLTACGTLTGIPAHGGKRFRAVEQELVAASSRAAVKEMDLSALKGRKAAL	50
		10 20 30 40 50 60	
5	orf83a.pep	YVSVMGDGSGNISGGRYSIDALIRGGYHNNPESATQYSYPAYDTTATTKSDALSVTTs	70 80 90 100 110 120
	orf83-1	YVSVMGDGSGNISGGRYSIDALIRGGYHNNPESATQYSYPAYDTTATTKSDALSVTTs	70 80 90 100 110 120
10	orf83a.pep	TSLLNAPAAALTKNSGRKGRERSAGLSVNGTGYRNETLLANPRDVSFLTNLIQTVFYLRG	130 140 150 160 170 180
	orf83-1	TSLLNAPAAALTKNSGRKGRERSAGLSVNGTGYRNETLLANPRDVSFLTNLIQTVFYLRG	130 140 150 160 170 180
15	orf83a.pep	IEVVPPPEYADTDVFVTVDFGTVRSRTELHLYNAETLKAQTKLEYFAVDRDSRKLLIAPK	190 200 210 220 230 240
	orf83-1	IEVVPPPEYADTDVFVTVDFGTVRSRTELHLYNAETLKAQTKLEYFAVDRDSRKLLIAPK	190 200 210 220 230 240
20	orf83a.pep	TAAYEYSQQYQYALWMPGPYSVGKTVKASDRLMVDFSDITPYGDTTAQNRPDFKQNNGKGP	250 260 270 280 290 300
	orf83-1	TAAYEYSQQYQYALWMPGPYSVGKTVKASDRLMVDFSDITPYGDTTAQNRPDFKQNNGKGP	250 260 270 280 290 300
25	orf83a.pep	DVGNEVIRRRKGKX	310
	orf83-1	DVGNEVIRRRKGKX	310
30	orf83a.pep		
	orf83-1		

35 Homology with a predicted ORF from *N.gonorrhoeae*

ORF83 shows 94.9% identity over a 197aa overlap with a predicted ORF (ORF83.ng) from *N. gonorrhoeae*:

	orf83.pep	TLLLFIPLVLTXCGLTGILAHGGKRFRAVEQELVAASSRAAVKEMDLSALKGRKAAX	58
40	orf83ng	MKTLLLLIPLVLTACGTLTGIPAHGGKRFRAVEQELVAASSRAAVKEMDLSALKGRKAAL	60
	orf83.pep	YVSVMGDGSGNISGGRYSIDALIRGGYHNNPESATQYSYPAYDTTATTKSDALSVTTs	118
	orf83ng	YVSVMGDGSGNISGGRYSIDALIRGGYHNNPDSATRYSPAYDTTATTKSDALSVTTs	120
45	orf83.pep	TSLLNAPAAALTKNSGRKGRERSAGLSVNGTGYRNETLLANPRDVSFLTNLIQTVFYLRG	178
	orf83ng	TSLLNAPAAALTKNSGRKGRERSAGLSVNGTGYRNETLLANPRDVSFLTNLIQTVFYLRG	180
50	orf83.pep	IEVVPPPEYADTDVFVTVDFGTVRSRTELHLYNAETLKAQTKLEYFAVDRDSRKLLIAPK	197
	orf83ng	IEVVPPPEYADTDVFVTVDFGTVRSRTELHLYNAETLKAQTKLEYFAVDRDSRKLLIAPK	240

The complete length ORF83ng nucleotide sequence <SEQ ID 317> is:

55	1	ATGAAAACCC TGCTCCTCT CATCCCCCTC GACTCTACCGG CCTGCCGCAC	
	51	ACTGACCGGG ATACCCGGCC ACGGCGGGGG CAAACGCTTT GCGCTCGAAC	
	101	AGGAACCTGG CGCCGCATCG TCCCGGGCCC CGCTCAAAGA AATGGACTTG	
	151	TCCGCCCTGA AAGAACCCCA AGCCGGCCCTA TACGTCTCCG TTATGGGGCA	
	201	CCAAGGTTGC GGCAACATAA CGCGCGGAGC CTACTCCATC GACCCACTGA	
	251	TACGGGGGGG CTACCATCAC RACCCGGCARA GGCGCACCCG ATACAGCTAC	
	301	CCGGCTCTAG ACAGTACCGCC CACCCACAAA TCCGACCGCC TCTCCGGCTT	
	351	RACCACTTC ACATGCTTGG TGAACGGCCC CGCCGCCGCC CTGACGAAATA	
	401	ACAAAGCGACG CAAAGCGGA CGCTCCGCCG GACTGTCCGT CAACGGCACG	
	451	GGCGACTTACG GCAACGAAC CCTGCTGCCG AACCCCGGG ACCTTTCCCT	
	501	CTCGACCAAC CTACATCAAA CGCTCTCTCA CCTGGCGGGG ATCGAAAGTCG	
	551	TACCGGGCGA ATACCGCGAC ACCGAGCTAT CGCTAACCGT CGACGTATTG	
65	601	GGCACCGTC GCAGCGTACG CGAACATGCC CTCTACACAG CGGAAACCCCT	

651 TAARGGCCAA ACCAAGCTCG ATAATTTCGC CGTOGACCGC GACAGCCGGA
 701 AACATGCTGAT TGCCCCCTAAA ACCGGCCGCTT ACGAATCCCC ATACCAAGAA
 751 CAATACGCCC TCTGATGATGCG ACCTTACAGC GTGCCAAAAA CGCTAACAGC
 801 CTCAGACGCC CTGATGATGCG ATTTCTTCGGA CATCACCCCC TACGGCGACCA
 851 CAACCGCCCA AAACCGTCCC GACTCTAACC AAAACAAAGG TAAAGAACCCC
 901 GATGTCGGCA ACGAAGTCAT CGCGCCGCGC AAAGGAGGAT AA

This encodes a protein having amino acid sequence <SEQ ID 318>:

1 MKTLLLLIPPL VLTAACGTLLTG IPAHHGGKRF AVEQELVAAS SRAAVKEMDL
 51 SALKGKRAAL YVSVMVGDGS GNISGGRYSI DALIRGVHNN NDSATRYSY
 101 PAYDTATTK SDALSGVTS TSLLNAAPAA LTKNNGRKGE RSQALSLNSGT
 151 GDYRNETLIAA NPROVSFLTN LIQTIVFYLRC IEVVPPEYAD TDVTVTVDVF
 201 GTVRSRTEHL LYNAETLKAQ TKLEYFAVDR DSRKLIIAPK TAATESQYQE
 251 QTALWGGPS VGNITVKASDR LMVDFSDITP YGDTTAQNRP DFKQNNGKNP
 301 DVGNEVIRRR KGS*

15 ORF83ng and ORF83-1 show 97.1% identity in 313 aa overlap

	10	20	30	40	50	60
orf83-1.pep	MKTLLLIPPLVLTAACGTLLTGIPAHHGGKRF	AVEQELVAAS SRAAVKEMDL	SALKGRKAAL			
orf83ng	MKTLLLIPPLVLTAACGTLLTGIPAHHGGKRF	AVEQELVAAS SRAAVKEMDL	SALKGRKAAL			
	10	20	30	40	50	60
20						
	70	80	90	100	110	120
orf83-1.pep	YVSVMMDQGSQNISGGRYSIDALIRGVHNN	NPPESATQISYPAYDTTATTKSDALSSVTTS				
orf83ng	YVSVMMDQGSQNISGGRYSIDALIRGVHNN	PDSATRYSYPAYDTTATTKSDALSGVTTS				
	70	80	90	100	110	120
25						
	130	140	150	160	170	180
orf83-1.pep	TSSLNAPAAALTKNRKGERSAGLSVN	GNTGDYRNETLLANPRDVSFLTNL	IQTIVFYLRG			
orf83ng	TSSLNAPAAALTKNRKGERSAGLSVN	GNTGDYRNETLLANPRDVSFLTNL	IQTIVFYLRG			
	130	140	150	160	170	180
30						
	190	200	210	220	230	240
orf83-1.pep	IEVVPPEYADTDVFVTVDVFGTVRSRTEL	LHYNAETLKAQTKLEYFAVDRDSRKLLITPK				
orf83ng	IEVVPPEYADTDVFVTVDVFGTVRSRTEL	LHYNAETLKAQTKLEYFAVDRDSRKLLIAPK				
	190	200	210	220	230	240
35						
	250	260	270	280	290	300
orf83-1.pep	TAAYESQTYEQYALNTGPFYKVSKTVKASDR	LMVDFSDITPYGDTTAQNRPDFKQNNNGKGP				
orf83ng	TAAYESQTYEQYALNMGGPSVGKTVKASDR	LMVDFSDITPYGDTTAQNRPDFKQNNNGKRN				
	250	260	270	280	290	300
40						
	310					
orf83-1.pep	DVGNEVIRRRKGX					
orf83ng	DVGNEVIRRRKGX					
	310					
45						
	50					
orf83-1.pep	DVGNEVIRRRKGX					
orf83ng	DVGNEVIRRRKGX					
	50					

Based on this analysis, including the presence of a putative ATP/GTP-binding site motif A (P-loop) in the gonococcal protein (double-underlined) and a putative prokaryotic membrane lipoprotein lipid attachment site (single-underlined), it is predicted that the proteins from *N.meningitidis* and 55 *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 38

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 319>:

```

5   1 ATGGCAGAGA TCTGTTTGT AACCGGCACG CCCGGTTCA GGGAAAACRTT
51 AAAATGTTT TCCATGATGG CGAATGATGA AATGTTTAAG CTCGATGAAA
101 101 AGCCCATACG CGTAGAAAGTA TTTACGAAAC TAAAGGCTT GAAAATACCG
151 151 CACACTTACA TAGAACCGA CGCAAAAAAG CTGCCGAAAT CGCACAGATGA
201 201 GCAGCTTCC CGCGATGATG TGACGATGAT GATAAAGAG CGCGAAAATA
251 251 TC GGCTCTAT TGTCAATTGTA GATGAGCTC ARGACCTATG GCGGGCACGC
301 301 TCGCAGGTT CAAAATCC TGAAAATGTC CAATGGCTGA ATACCCACAG
351 351 ACATCAGGGC ATTGATGTTAT TTGTTTTGAC TCAAGGTTCT AAGCTTCTAG
401 401 ATCAAATCT TAGAACGCT GTACCGAAC ATTACCATC CGCTTCAAC
451 451 AAAGTGGTA TCCTGATCGCT TTGAGATGG AAAATATGG CGGACGATCC
501 501 CGTAAATAGT GCATCAGGGC CATTCCTCCAG TATCTTACAG TATCTATACA CTGATATAAA
551 551 AAAGTTTATG CTGCTAysx GGGGGGAGG TTCTATACCTT AATAAAGGTC
601 601 AGACGGCTCA AGTGTTTGTG CACTCTGCc GTAAATGATG TGCTGATTC
651 651 CGTGTGTTGTC GGCGCTGCTCT ATAAATAGTT GagcGTTAC GGGAAAAAAC
701 701 aGGAGAACCC CGCGACACA GAATCGGGG CAACAGAACA CGAGGCACTA
751 751 CTTCCGGATA AACAGAAGG CGGGCGGTG AATRACCGCA ACCTTACCGC
801 801 AGATATGTTT GTTCCGACAT TGTCGCSAAA ACCCGAGACG AACCGAGATT
851 851 ATAACGTTG AAGGGAGGTG AGAACCTTG AATATATAGC AGGCTGTATA
901 901 GAGGCGGAA GCACCGGATG CGCCGCTGAT TCAGGAGAAC GGGACGCGATC
951 951 ggAAAGAGT ACGGAGGTG TGTCGcgAg cCTATGtAA AAcgGCTTG
1001 1001 CGTTTAAAC CATAAGAAGA AGAAAGCCAA GGGCAAGGAG TTGACCAAAG
1051 1051 CGCGCAGCA CATTCGGACA GGCGCGAACG TTGCGCACAT GGGCGAAA
1101 1101 CGCTAGAGA ACCTAATGTA CGATAATTGG AGAACAGGG GGGAAACGTT
1151 1151 TGAGGAATG CGGGGGC GTGTCGGAT CGCGAAAATC A

```

This corresponds to the amino acid sequence <SEQ ID 320; ORF84>:

```

30   1 MAEICLITGT PGSGCTTLKVN SNNANDEMPK PDEKAIRRKV FTNIKGLKIP
51 HTYIETDAKK LPKSTDEQLS AHDMYEWIKK PENIGSIVIV DEAQDWWMPAR
101 101 SAGSKIPEVN QWLNTHRHQS IDIEFLVTOGE KLDQNLIRTL VRKHVHIAHN
151 151 KMGMRLLWEK KICADPPVKM ASSAFSSIIYT LDKKVVVDLYX XAEVHTVNKV
201 201 KRSKWFTYLP VIVLILPVVF GLSYKMLLSY GKKQEEPAAQ ESAATEQQAV
251 251 LPDKEFEGPV NNGNLTDAMT VPTLSEKFPX KPIYNGVRQV RTFEYIAGCI
301 301 EGGRITGACY SHQGTAALKEV TELMRCKDYVR NLGFNPYKE BSQGQEVEQVS
351 351 AQQHSRSRAQV ATLGGKPxQN LMNDWEERGK KPFEGIGGGV VGSAN*

```

Further work revealed the complete nucleotide sequence <SEQ ID 321>:

```

40   1 ATGGCAGAGA TCTGTTTGT AACCGGCACG CCCGGTTCA GGGAAAACATT
51 AAAATGTTT TCCATGATGG CGAATGATGA AATGTTTAAG CTCGATGAAA
101 101 AGCCCATACG CGTAGAAAGTA TTTACGAAAC TAAAGGCTT GAAAATACCG
151 151 CACACTTACA TAGAACCGA CGCAAAAAAG CTGCCGAAAT CGCACAGATGA
201 201 GCAGCTTCC CGCGATGATG TGACGATGAT GATAAAGAG CGCGAAAATA
251 251 TC GGCTCTAT TGTCAATTGTA GATGAGCTC ARGACCTATG GCGGGCACGC
301 301 TCGCAGGTT CAAAATCC TGAAAATGTC CAATGGCTGA ATACCCACAG
351 351 ACATCAGGGC ATTGATGTTAT TTGTTTTGAC TCAAGGTTCT AAGCTTCTAG
401 401 ATCAAATCT TAGAACGCT GTACCGAAC ATTACCATC CGCTTCAAC
451 451 AAAGTGGTA TCCTGATCGCT TTGAGATGG AAAATATGG CGGACGATCC
501 501 CGTAAATAGT GCATCAGGGC CATTCCTCCAG TATCTTACAG TATCTATACA CTGATATAAA
551 551 AAAGTTTATG CTGCTAysx GCGCGGAAG TTCTATACCTT AATAAAGGTC
601 601 AGCGGTCAA AGTGTTTGTG CACTCTGCc GTAAATGATG TGCTGATTC
651 651 CGTGTGTTGTC GGCGCTGCTCT ATAAATAGTT GAGCAGTTAC GGGAAAAAAC
701 701 AGGAAGAACCC CGCGACACA GAATCGGGG CAACAGAACA CGAGGCACTA
751 751 CTTCCGGATA AACAGAAGA CGGGCGGTG AATACCGCA ACCTTACCGC
801 801 AGATATGTTT GTTCCGACAT TGTCGAAAAA ACCCGAAAGC AGGCGGATT
851 851 ATAACGTTG AAGGGAGGTG AGAACCTTG AATATATAGC AGGCTGTATA
901 901 GAGGCGGAA GCACCGGATG CGCCGCTGAT TCAGGAGAAC GGGACGCGATT
951 951 GGAGAACAGT CGGGAGGTG TGTCGAGGA CTATGTAATAA ACAGGCTTGC
1001 1001 CGTTTAAAC CATAAGAAGA AGAAAGCCAA GGGCAAGGAGT TCAGGAAAGC
1051 1051 CGCGCAGCAAC TTGGCGACAG GGGCGAACGT GCACATTTG GGGGAAAAGC
1101 1101 GTAGCAGAAC CTAATGTAAG ATAATTGGGA AGAACGGGGG AAACCGTTTG
1151 1151 AGGAATCTGG CGGGGGCTGTG GTGCGATCGG CAAACTGTA

```

This corresponds to the amino acid sequence <SEQ ID 322; ORF84-1>:

```

1 MAEICLITGT PGSGKTLKVM SMMANDEMFK PDENGIRRKV FTNIKGLKIP
51 HTYIETDAKK LPKSTDEQLS AHDMYEWIKK PENIGSIVT DEAQDWPAR
101 SAGSKIPENV QWLNTHRHCG IDIFVLTOGP KLDQNLRL VRKHYHIASN
151 KMGMTLLEW KICADDPVKM ASSAFSISYT LDKKVVYDLYTE SAEVHTVNKV
5 201 KRSKWFTTLP VIVLILIPFV GLSYKMLSSY GKKEEPAQQ ESSATEQQAV
251 LPDKTGEPEV NGNLTAIDM VFPTLSEKPE KPITYNGVRQV RTFEYIAGCI
301 EGGRGTCACY SHQGTALKEV TELMCCKDYVK NGLFPNPYKE ESSGQEVCQGS
351 AQQHSDRAQV ATLGGKP*QN LYMDNWEEERG KPFEGIGGGV VGSAN*

```

Computer analysis of this amino acid sequence gave the following results:

10 **Homology with a predicted ORF from *N. meningitidis* (strain A)**

ORF84 shows 93.9% identity over a 395aa overlap with an ORF (ORF84a) from strain A of *N. meningitidis*:

	10	20	30	40	50	60
15 orf84.pep	MAEICLITGT PGSGKTLKVM SMMANDEMFK PDENGIRRKV FTNIKGLKIP					
orf84a	MAEICLITGT PGSGKTLKVM SMMANDEMFK PDENGIRRKV FTNIKGLKIP					
20 orf84.pep	10	20	30	40	50	60
	LPKSTDEQLSAHDMYEWIKK PENIGSIVT DEAQDWPAR SAGSKIPENV QWLNTHRHCG	70	80	90	100	110
orf84a	10	20	30	40	50	60
	LPKSTDEQLSAHDMYEWIKK PENIGSIVT DEAQDWPAR SAGSKIPENV QWLNTHRHCG	70	80	90	100	110
25 orf84.pep	130	140	150	160	170	180
	ID1FVLTQCGPKSLDQNLRLVRKHYHIASNKGMRNLTLEWK1CADDPVKMASSAFSS1YT	130	140	150	160	170
orf84a	130	140	150	160	170	180
	ID1FVLTQCGSKLQLDQNLRLVRKHYHIASNKGMRNLTLEWK1CADDPVKMASSAFSS1YT	130	140	150	160	170
30 orf84.pep	190	200	210	220	230	240
	LDDKKVYDLYXXAEVHTVNKVKRSKWFYTLPVIVLILPVVFVGLSYKMLSSYGGKQEEPAQQ	190	200	210	220	230
orf84a	190	200	210	220	230	240
	LDDKKVYDLYESAEVHTVNKVKRSKWFYTLPVIVLILPVVFVGLSYKMLSSYGGKQEEPAQQ	190	200	210	220	230
35 orf84.pep	250	260	270	280	290	300
	ESAAATEQQAALPDKTEGEPVNNNGNLTAIDMVFPTLSEKPKSKP1IYNGVRQVRTFEYIAGCI	250	260	270	280	290
orf84a	250	260	270	280	290	300
	ESAAATEHQAVFQDKTEGEPVNNNGNLTAIDMVFPTLSEKPKSKP1IYNGVRQVRTFEYIAGCV	250	260	270	280	290
40 orf84.pep	310	320	330	340	350	360
	EGGRGTCACYSHQGTALKEVTELCKDYYVNGLPNPYKEESOCOEVQOSAOGHSRDRQV	310	320	330	340	350
orf84a	310	320	330	340	350	360
	EGGRGTCCTCYSHQGTALKEVTELCKDYYVNGLPNPYKEESOCOEVQOSAOGHSRDRQV	310	320	330	340	350
45 orf84.pep	370	380	390			
	ATLGGKPXQNLMYDNWEERGKPFEGIGGGVGVSANK	370	380	390		
orf84a	370	380	390			
	ATLGGKPFWQNLMYDNWEERGKPFEGIGGGVGVSANK	370	380	390		
50 orf84.pep						
55 orf84.pep						
60 orf84a	1 ATGGCAGAGA TCTGTTTGAT ACCGGCACG CCCGGTTTCAG GGAAAACATT					
	51 AAAATGGTT TCCATGATGG CAACAGTGA AATCTTAAAG CGCGATGAAA					
101 ACGGCATACG CCGTAGAAGTA TTTAGGACCA TCAAAGGCTT GAAAGATCCG						
	151 CACACTTACA TAGAACCGA CGCGAAAAG CTGCCGAATT GCACAGATGA					
201 CGACGCTTCG GCGCATGAT TGTCAGGATG GATAAGAAG CCGGAAAATA						
	251 TCGGGTTCTG TTGTCAGTGA GATGAAAGCTT AGAGCGTATG GCGGGCAGC					
301 TCGGGAGGTT CAAAAATTCG CAATGGATTCG CAATGGCTGA ATACGCACAG						
	351 ACATCAAGGC ATTGATCATC TTGAAAGATC CAAAGGCTCT AAGCTTCTAG					
401 ATCAAAATCT TAGAACGCTT GTACGGACAC ATTACCACT CGCTTCAAAC						
	451 AAGATGGTA TCGCTACGCT TTTGAATGCG AAAATATGGG CGGACGATCC					

The complete length ORF84a nucleotide sequence <SEQ ID 323> is:

```

55
1 ATGGCAGAGA TCTGTTTGAT ACCGGCACG CCCGGTTTCAG GGAAAACATT
51 AAAATGGTT TCCATGATGG CAACAGTGA AATCTTAAAG CGCGATGAAA
101 ACGGCATACG CCGTAGAAGTA TTTAGGACCA TCAAAGGCTT GAAAGATCCG
151 CACACTTACA TAGAACCGA CGCGAAAAG CTGCCGAATT GCACAGATGA
201 CGACGCTTCG GCGCATGAT TGTCAGGATG GATAAGAAG CCGGAAAATA
251 TCGGGTTCTG TTGTCAGTGA GATGAAAGCTT AGAGCGTATG GCGGGCAGC
301 TCGGGAGGTT CAAAAATTCG CAATGGATTCG CAATGGCTGA ATACGCACAG
351 ACATCAAGGC ATTGATCATC TTGAAAGATC CAAAGGCTCT AAGCTTCTAG
401 ATCAAAATCT TAGAACGCTT GTACGGACAC ATTACCACT CGCTTCAAAC
451 AAGATGGTA TCGCTACGCT TTTGAATGCG AAAATATGGG CGGACGATCC

```

5 501 CGTAAAAATG GCATCAAGCG CATTCTCCAG TATCTATACA CTGGATAAAA
 551 AAGTTTATGA CTGTAGCAA TCACGGAAAG TTCACTACCGT AAATAAAGGTG
 601 AAAGGGTCAA ATATGGTTTA TACTCTGCCA GTATAATAT TGCTGATITC
 651 CGTTTTGTC GCCTCTGCCT ATAAAATGTT AAAGTAGTTAT GGAAAAAAAC
 701 AGGAAGAACCC CGCACACAA GAATCGCGGC CAACAGNACA TCAGGCAGTA
 751 TTTCAGGATA AAACAGAAGG TGAGCCGGTG AACAAACGGTA ACCCTTACCGC
 801 AGATATGTTT GTTCCGCAAT TGTCGAAAAC ACCCGAAAGC AAAGCCGATTT
 851 ATAACCGGTTG AAAGCAGGTA AGAACCTTIG ATATATATAGC AGGCTGTGTYA
 901 GAAGGGGAA GRACCGGATG CACATCTAT TGCGATCAAG GGACGGCATT
 951 GAAAGAATT ACAAAAGGAA TGTCAGGAA TTACCCAAGA AACCGGATTGC
 10 1001 CGTTTACCC ATATAAAGAA GAACGCAAG GGGGGGATGT CCAGCAAAGT
 1051 GAGCAGACACC ATTCCGACAG ACCCGAAGT GCCACGTTGG GCGGAAAGCC
 1101 GTGCAAAAT CTATGATAG ATAATGGCA CGAGCCGGGA AAACCGTTTG
 1151 AAGGAATCG CGGGGGCGT GTCGATCGG CAAACTGA

15 This encodes a protein having amino acid sequence <SEQ ID 324>:

1 MAEICLITGT PGSGKTLKVM SMMANDEMFK PDENGIRRKY FTNIKGLKIP
 51 HTYIETDAKK LPKSTDEQLS AHDMYEWIKK PENIGSIVTV DEAQDWPAR
 101 SAGSKIPENV OWLNTHRHOG IDIFVLTGGS KLLDQNLRLT VRKHYHIAHN
 151 KMGMRTLLEW KICADDPVKK ASSAFSSIIYT LDKKVYDLYE SAEVHTVNKV
 20 201 KRSKWFYTLPL VILLILPVFV GLSKYMLSSY GKQKEEPAAO ESAATEHQAV
 251 FQDTTEGEPVV NNGNLTDADMF VFTLSEKPKES KPIYNGVRQ RTFEYIAGCV
 301 EGGRGTCTCY SHQGTALKWEI TKEMCKDYAR NGLPPNPYKE ESQGRDVQQS
 351 EQHISDRPQV ATLGGKPWNL MYDNWQERGKPFEGIGGGVVGSANX

ORF84a and ORF84-1 show 95.2% identity in 395 aa overlap:

	10 20 30 40 50 60
25 orf84a.pep	MAEICLITGT PGSGKTLKVM SMMANDEMFK PDENGIRRKY FTNIKGLKIPHTYIETDAKK
	10 20 30 40 50 60
orf84-1	MAEICLITGT PGSGKTLKVM SMMANDEMFK PDENGIRRKY FTNIKGLKIPHTYIETDAKK
	70 80 90 100 110 120
30 orf84a.pep	LPKSTDEQLSAHDMYEWIKKPEPINIGSIVTVDEAQDWPARSAGSKIPENVQWLNTHRHG
	70 80 90 100 110 120
orf84-1	LPKSTDEQLSAHDMYEWIKKPEPINIGSIVTVDEAQDWPARSAGSKIPENVQWLNTHRHG
	130 140 150 160 170 180
35 orf84a.pep	IDIFVLTQGSKLLDQNLRLTVRKHYHIAHNKMGMRTLLEWKICADDPVKMASSAFSSIIYT
	130 140 150 160 170 180
orf84-1	IDIFVLTQGPKLLDQNLRLTVRKHYHIAHNKMGMRTLLEWKICADDPVKMASSAFSSIIYT
	190 200 210 220 230 240
45 orf84a.pep	LDDKKVYDLYESAEVHTVNKVRSKRWYTTLPVIIILIPFVGLSYKMLSYGKQKEEPAAO
	190 200 210 220 230 240
orf84-1	LDDKKVYDLYESAEVHTVNKVRSKRWYTTLPVIIILIPFVGLSYKMLSYGKQKEEPAAO
	250 260 270 280 290 300
50 orf84a.pep	ESAAATEHQAVFQDKTTEGEPVNNGNLTA DMFVPTLSEKPKESKPIYNGVRQVRTFEIAGCV
	250 260 270 280 290 300
orf84-1	ESAAATEQQAVLPDKTTEGEPVNNGNLTA DMFVPTLSEKPKESKPIYNGVRQVRTFEIAGCI
	310 320 330 340 350 360
55 orf84a.pep	EGGRTGCTCYSHQGTALKWEITKEMCKDYARNGLFPNPyKEESQGRDVQQSEQHHSRDPQV
	310 320 330 340 350 360
orf84-1	EGGRTGCTCYSHQGTALKWEITLMECKDYARNGLFPNPyKEESQGRDVQQSEQHHSRDPQV
	370 380 390
60 orf84a.pep	ATLGGKPWNL MYDNWQERGKPFEGIGGGVVGSANX
	370 380 390
orf84-1	ATLGGKPWNL MYDNWQERGKPFEGIGGGVVGSANX

Homology with a predicted ORF from *N.gonorrhoeae*

ORF84 shows 94.2% identity over a 395aa overlap with a predicted ORF (ORF84.ng) from *N. gonorrhoeae*:

5	orf84.pep	MAEICLITGTGPGSCKTLKMSVSMMANDEMFKPDEKAIRRKVFTNIKGLKIPHTYIETDAKK	60
	orf84ng	MAEICLITGTGPGSCKTLKMSVSMMANDEMFKPDEKAIRRKVFTNIKGLKIPHTYIETDAKK	60
10	orf84.pep	LPKSTDEQLSAHDMYEIKKPENIGSIVIVDEAQDVWPARSAGSKIPEVNQWLNTTHRHQG	120
	orf84ng	LPKSTDEQLSAHDMYEIKKPENIGSIVIVDEAQDVWPARSAGSKIPEVNQWLNTTHRHQG	120
15	orf84.pep	IDIFVLTQGPKLLDQNLRVLVRKHYYHIAANRKGMLRTLLEWKVCAADDPVKMASSAFSSIYT	180
	orf84ng	IDIFVLTQGPKLLDQNLRVLVRKHYYHIAANRKGMLRTLLEWKVCAADDPVKMASSAFSSIYT	180
20	orf84.pep	LDKVYVLYXXAEVHTVNKVKRSKWFYTLPLVIVLILPVFVGLSYKMLSSYGKQEEPAQ	240
	orf84ng	LDKVYVLYXXAEVHTVNKVKRSKWFYALPVVILLIPFVGLSYKMLSSYGKQEEPAQ	240
25	orf84.pep	ESAAATEQQAVLPDTEGEFPVNNGNLTADMVFPTLSEKEPKSKPFIYNGVRQVRTFEYIAGCI	300
	orf84ng	ESAAATEQQAVLPDTEGEFPVNNGNLTADMVFPTLPEPKESPKIYNGVRQVRTFEYIAGCI	300
30	orf84.pep	EGGRTGCACYSHQGTALKETVTELCKDYYVNGLPFPNPKYKEESQQEVQOSAQOHSDRAQV	360
	orf84ng	EGGRTGCACYSHQGTALKETVTELCKDYYVNGLPFPNPKYKEESQQEVQOSAQOHSDRAQV	360
35	orf84.pep	ATLGKPKXPNQMLYDNWEERGKFFEGIGGGVVGSAN	395
	orf84ng	ATLGKPKXPNQMLYDNWEERGKFFEGIGGGVVGSAN	395

The complete length ORF84ng nucleotide sequence <SEQ ID 325> is:

1	ATGGCAGAAA	TCTCTTTGAT	AACCGGCCAC	CCCGTTTCA	GGAAAAACATT
51	AAAAATGGTT	TCCATGATGG	CAAACGATGA	AATGTTTAAG	CCAGATGAAA
101	ACGGGTCACC	CCTGAACTG	TATTCAGAAC	TCAAAGGTTT	GAAGATACCG
151	CACACCCACA	TAGAACACAGA	CSCMAAGAAC	CTGCCAAAT	CAACCGATGA
201	ACAGCTTCCG	GCGGATGATA	TGTTATGATG	GATCAAGAAAG	CCTGAAAacg
251	tcggcgAAAT	CSTTATTGTC	GATGAGGCCG	AAAGACTATG	GCCCCGACGC
301	TccgCAGGTT	CGAAAATCCC	CGAACAACTG	CAATGGCTGA	ACACACACAG
351	GCATCACGGC	ATAGATAGAT	TCTGATGAC	ACAAGGTCCT	AAACCTTCTAG
401	ATCAGAACCT	GGCGAACATG	GTTAAAAGAC	ATTACACCAT	TGGGGCCAAC
451	AAAATGGTT	TGGCTACCTT	GCTTGATGGT	AAAGTATGGG	CGGATGACCC
501	GGTAAATAG	GCATCACATG	CATTTCCTAG	TATCTACACA	CTGGATAAAA
551	AAAGTTATGA	CTTGATGAA	TCCGCGAAAA	TTCACACGG	AAACAAACTC
601	AAGCGCTCAA	AATGGTTTA	TGCAATGCC	GTCATCATAT	TATGATTC
651	GCTTATGTC	GGTTTGCTCT	ACAAAATGTT	GGCGAGTTAC	GGAAAAAAAC
701	AGGAAGAAC	CGCGCACAG	GAATCGCGG	CAACAGAAC	GCAGGAGTA
751	CTTCGGATA	GGAAAGAACG	GAATAGCGT	ATAACCGGA	ACCTTACCGC
801	AGATATGTTT	GTTCGGCAT	TGCGCGAAA	ACCCGAAAGC	AGCGGATT
851	ATAACGGTGT	AAAGCGCTGA	AGGACCTTGT	ATAATATAGC	AGCGTGTATA
901	GAAGGCGAA	GAACCGGAT	CACCTGTAT	TGCGCATCA	GGACGGCATT
951	GAAGAACAGT	ACGGAGTGTG	TGTCGAAGGA	CTATGTTAAA	AAAGGCTTGC
1001	CGTTAACCC	ATTCGGACAC	GGCGCAAGTT	GGCACCTTGG	GGCGAAAACC
1051	GGCCAGCAAC	ATTGGGACAC	GGCGCAAGTT	GGCACCTTGG	GGCGAAAACC
1101	GCAGCAGAAC	CTAATGTACG	ACAAATGGGA	AGAACGGGG	AAACCGTTG
1151	AAGGAATCGG	CGGGGGCGTG	GTGCGATCGG	CAAACGTGA	

This encodes a protein having amino acid sequence <SEQ ID 326>:

1	MAEICLITGT	PGSGSCKTLKMV	SMMANDENFK	PDENGVRRKV	FTNIKGLKIP
51	HTHIEIDAKK	LFRSTDEQLS	AHDHYEWIKK	FENVGAIIVIV	DEAQDVWPAR
101	SASGKIPENV	QWLNTHRHOG	IDIFVLTQGP	KLDQNLRTL	VKRHYHIAAN
151	KMCLTLLKEW	VKCADDPVK	ASSAFSSIYT	LDDKKVYDLYE	SAAIHTVNKV
201	TKRSKWFYALP	VILLILPVF	GLSYKMLSSY	GKQEEPEAQ	ESAAATEQQAV
251	LDPKTEGEVS	NNGNLTDME	VPTLPEPKES	KPIYNGVRQV	RTEYIAGCI
301	EGGRTGCACY	SHQGTALKET	TELCKDYYV	NGLFPNFYKE	ESGGQEVQQS
351	AQHQHSDRAQV	ATLGKPKQON	IMYDNWEERG	KPFEGIGGGV	VGSAN*

ORF84ng and ORF84-1 show 95.4% identity in 395 aa overlap:

		10	20	30	40	50	60
5	orf84-1.pep	MAEICLITGTPGSGTKLKVMVSMMANDEMFKPDENGIRRKVFTNIKGLKIPHTYLETDAAKK					
	orf84ng	MAEICLITGTPGSGTKLKVMVSMMANDEMFKPDENGIRRKVFTNIKGLKIPHTYLETDAAKK	10	20	30	40	50
10	orf84-1.pep						
	orf84ng	LPKSTDEQLSAHDMDYEWIKKPENIGSIVIVDEAQDVWPARSGASKIPEVNQWLNTHRHOG	70	80	90	100	110
15	orf84-1.pep						
	orf84ng	LPKSTDEQLSAHDMDYEWIKKPENIGSIVIVDEAQDVWPARSGASKIPEVNQWLNTHRHOG	70	80	90	100	110
20	orf84-1.pep						
	orf84ng	IDIFVLTQGPKLLDONLRLTVRKHYHIAANKGMRLLWEWKVCAADDPVKMASSAFSSIYT	130	140	150	160	170
25	orf84-1.pep						
	orf84ng	IDIFVLTQGPKLLDONLRLTVRKHYHIAANKGMRLLWEWKVCAADDPVKMASSAFSSIYT	130	140	150	160	170
30	orf84-1.pep						
	orf84ng	LOKKVYDLYESAEVTEVNKVKRSKWVFTEPLVIVLPLVFGSLSYKMLSSYGKKQEPEAAQ	190	200	210	220	230
35	orf84-1.pep						
	orf84ng	ESAAATEQQAVLPDKTEGEFVNNGNLTADMFVPTLSEPKESKPIYNGVRQVRTFEYIAGCI	250	260	270	280	290
40	orf84-1.pep						
	orf84ng	ESAAATEQQAVLPDKTEGEFVNNGNLTADMFVPTLSEPKESKPIYNGVRQVRTFEYIAGCI	250	260	270	280	290
45	orf84-1.pep						
	orf84ng	EGGRRTGCACYSHQGTALKVTELMCKDYVKNGLPPNFYKEESQGQEVOQSAQQHSdraov	310	320	330	340	350
50	orf84-1.pep						
	orf84ng	EGGRRTGCCTCYSHQGTALKVTELMCKDYVKNGLPPNFYKEESQGQEVOQSAQQHSdraov	310	320	330	340	350
55	orf84-1.pep						
	orf84ng	ATLGGKPKQNLMYDNWEERGKPFEGIIGGGVGUVGSANX	370	380	390		
60	orf84-1.pep						
	orf84ng	ATLGGKPKQNLMYDNWEERGKPFEGIIGGGVGUVGSANX	370	380	390		

Based on this analysis, including the presence of a putative transmembrane domain (single-underlined) in the gonococcal protein, and a putative ATP/GTP-binding site motif A (P-loop, double-underlined), it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 39

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 327>:

1	GTTGGTTTC	TGAATGCCGA	CAACGGGATA	TTGGTTCAAGG	ACTTGGCCTT
51	TGAAGTCAAA	CTGAAAAAAT	TCCATATCGA	TTTTTACAAAT	ACGGGTATGC
101	CGCGTGATT	CGCCAGCGAT	ATTGAAGTGA	CGGACAAGGC	AACCGGTGAG
151	AAACTCGAGC	GCACCATCG	CGTGAAACAT	CCTTTCGACCT	TGACACGCAT
201	CACGATTAT	CAGGGAGATT	TTGCCGACGG	CGGTTCGCGAT	TTGACATTCGA
251	AGGGCTGGAA	TTGGGTGAT	GCTTCGCGC	AGCCCTGCGT	TTGAAGGCA
301	ACATCCATAC	ACCGATTTCC	GTGGAAATGT	GGCAAACACAA	AATATGCTCT
351	TGAGTCGGAT	CAGTTCACATT	CTATGAGATG	GGAGGACATG	AGCGAGGGCG
401	CGGAAGCGGG	AAAAGACCTG	AAATCCACGC	TGCGCCGATGT	CGCGGCGCTT
451	ACTCAGGAGA	GTCACAAATA	CACCAAT...TACCG
501	TATCCGTGAT	GGCCGAGGCC	AGGGGTGCGA	ATATAAAAAC	TATATGCTGC
551	CGGTTTGCA	GGAACAGGGAT	TATTTTGGA	TTACCGGCCAC	CGCGACCGC.

601 TTGCGACGAG AATACGGCTG GTGTGGTATC CCCTTGGACA AGCAGTTGAA
 651 AGCGGACACC TTTATGGCAT TCGCTGAGTT TTGAAAAGAT GGGGAAGGGC
 701 CCAACCGT T.GTTCGGCAC GCAACCAAG GCGCACCTGC CGAAATCGC
 751 GAACAACTTA TGCTGGCTG GGAAAACACG CTGACATCTT TTGCAACAAA
 5 801 AGGCTATTG GGATGGAC AATTATTAC CCTCAATATC CGCAAAGGC
 851 AGCGAGATAA GATCGAGGC TATTCTACG AAATGCTTAA CGCCGTGATG
 901 AACGCTCTT TGGATSAAC CAT ACCGGG TACGCGCTTC CGCAATGCCA
 951 CGAGGATGAA CGCGGGAATC GTTCTCTGCT GCACAGTATG GTAGGGTACA
 1001 CGGGTTGAC CGAATATCC CGGCTATGC TGCTGCAACT TGATGGTT
 10 1051 TCCGAGGTGC GTTCGTCGGG TTTCAGATG ACCCGTCCC C.GGTCCCCT
 1101 TTGGCTAT CTC...

This corresponds to the amino acid sequence <SEQ ID 328; ORF88>:

1 MVFLNADNGI LVQDLPFEVK LKKEFHIDFYN TGMPRFASD IEVTDKATGE
 5 KLERITVRNH PLTLHGLITIY QASFADGGSD LTFFKANWLD ASREFPVVLKA
 15 101 TSIIHOFELI GKHHYRLEED QFTSMVNEDN SEGAEREKSL KSTLPDVRAV
 151 TQEKGHXYTNX TQQXXYXRIRI AFQGQAVEYKN YMLPVLIQED YFWIITGTSX
 201 LQQQYEWIRI PLDKQLKADT FMALEFILKD GEGRKRXXVAD ATKGAPAEIR
 251 EQFMQLAETN UNIFIAQKGYL GLDEFITSNI PKEQODKMGY YFYEMLYGV
 301 NAAEDETDXTR YGLFEEWQDDE ARNRFLLHSR DAYTGLTEYP APMLLQLDF
 20 351 SEVRSSGLQM TRSXGLLVLV Y...

Further work revealed the complete nucleotide sequence <SEQ ID 329>:

1 ATGAGTAAAT CCCGTAGATC TCCCCCACTT CTTTCCCGTC CGTGGTTGCG
 51 TTTTTCAGC TCCATCGCT GTTCACTGTCG TTTCAGTCGC TTGCTGGGTA
 101 TTGCACTCGT TATCGCTACG GTTTCAGCAGG AAAACCAAGCC GCAGACGGAT
 151 TATTGGTCA AATTCCGATC GTTTGGCCG CAGATTTTG GTTTCTGGG
 201 ACTGTAGACG GTCTGATCG CGCATGGTT TGCTGTTATC ATGATGTTTT
 251 TTGGTGGTTT TACAGTTTG TGCTGATTC CGAATGTCGC CGCGTCTGG
 301 CGCGGAATGA AGCTCTTTGC GGAAAAGGTT AAAGAAAAAAT CTTGGCGGC
 351 GATGGCCGATC TTCTTCGTCG TGATGTAAN AATTGGCCC GAGGTGCGA
 401 AACGTATCTT GGAAGTACAA GGTTTCTACG GAAAACCAT TAACCGTGAA
 451 GACGGGCTGG TTCTGGATTC CTTTGATTGT CATTGCTCTG GGCGGGTTGA
 501 CTATATCTT CCCCCATGTT CCTCTGGTTT AAACCTGGTA TGCTGACCGS TCGATGTT
 551 TAGACAGTAA CCTCTGGTTT CTCAGCTTTA TGCAAGGGAT TTCAAGGGCA AAAGTATTTT
 601 CGGGCAACATC AGGGCGCTTC TGCAAGGGAT TTCAAGGGCA AAAGTATTTT
 651 GGGTGGCTTC AATCTCTCAT TTAGGGGCA CGTCAATATT TCCGAGGGC
 701 AGAGTGGCGA TTGTGGTTTC CTGAATGCCA ACAACGGGAT ATTGGTTCAG
 751 GACTTGGCTT TTGAACTGCA CTGAAAGAAA TTCCATATCG ATTTCATCAA
 801 TACGGTATG CGCGCTGATI TGCGCACCGGA TTATGGAATG ACGGACAAAGG
 851 CAACGGTGA GAAACCTGC CGCACCCATCC CGCTGAAACCA TCCCTTGAC
 901 TTGCAAGCGCA TCAGCTTAA TGAGGGATTT TTGCGGAGC GCGGTTCCGA
 951 TTGCACTTC AAGGGCGTGG AATTGGGTA TGCTTGGCC GAGCTCTGCG
 1001 TTGTAAGGC AACATCCATA CACCAAGTTT CGTGGAAAT TGCAAAACAC
 1051 AAATATCGT TTGAGTGGCA TCGATTCCTC TCTATGAATG TGGAGGACAT
 1101 GAGCGAGGGC CGCGAACGG AAAAAGGCTT GAAATCCACG CTGAACATG
 1151 CGCGGAAGAG TACTCGAGA GTGAAAAAAAT ACACCAAATAT CGGCCCCCTCC
 1201 ATTGTTTACG CTATCGCTG TTGGCCAGGG CAGGCGTTGG AATATAAAA
 1251 CTATATGCTG CGGTGTTTG AGGAAACAGGA TTATTTTTGG ATTACCGGCA
 1301 CGCGCACCGG CTTCAGCAC CGAACCCGTC CGTGGCTAT CCCTCTGGAC
 1351 AAAGCTGTA CGACGGACAC TCAATTCGCA TTGGCTGAGT TTGTTGAAAAGA
 1401 TGGGGAGGGC CGCAACGTC TGTTGGCGA CGCAACCAA GGGCACCTG
 1451 CGCAAACTTC CGAACAAATC ATGCTGGCTC CGGAAACAC GCTGAACATC
 1501 TTGGCACAAA AAGGCTTATG GGAAATTGGAC GAAATTATTA CGCCCAATAT
 1551 CGCGGAAGAG CAGCAAGATA AGATGCAAGG CTATTCATC GAAATGCTTT
 1601 ACGGCGTGTGAA GAGCGCTGT TTGGATGAAA CGATAGCGG GTACCGCTG
 1651 CGCGGAATGCA AGCAGGTGAA AGCGCGGAAT CGTTTCTCTG TGCAACAGTAT
 1701 GGRTGCTGAC CGCGTTTGAA CGGAATATCC CGCCCTCTATG CTGCTGCAAC
 1751 TTGATGGTTT TTCCGGAGTG CGTTCGTCGG GTTGGAGAT GACCGGTTCC
 1801 CGGGTGGCGC TTTCGGCTTA TCCTCGCTCG GTGCTGTTGG TATTTGGTAC
 1851 GGTTATGTC TTATATGTC GCGAAAACG GGCCTGGGTG TTGTTTCAG
 1901 ACGGCAAAAT CGGTCTGGC ATGTCCTGGG CCCCGACGCA ACGGGATTTC
 1951 CAGGAAGAT TTCCAAACAA CGTCGAGAGT CTGCAACGSC TCAGCAAGGA
 2001 CTGAAATCAT GACTGA

This corresponds to the amino acid sequence <SEQ ID 330; ORF88-1>:

1 MSKSRSPPL LSRPWFAFFS SMRFAVALLS LLGIASVIGT VLOQNQPOTD
 65 51 YLVFKGSFWA QIFGFLGLYD VYASAWVVI MMFLVVSTSL CIRNVPPFW

101 REMKSFREKV KEKSLAAMRH SSILOVKIAP EVAKRYLEVQ GFQGKTFINRE
 151 DGSVLLIAARKK GITMKNGYIF AHVALIVLICL GGLLIDSNLII KLGMLTGRIV
 201 PDNQAVAYAKD FRPEPSILGK NLSFRGNVNTI SEQQSADVVFI LNADNGILVQ
 251 DLPEVFKLKDFIDFYNTGM PRDFASDIEV TDKATGEKLE RTTRVNHLPT
 5 301 LHGITYQAS FADGGSDLTFC KAWNLDASR EPWVILKATSI HQFFPLEIGKH
 351 KYRLEDQFT SMNVEDMSSEG AEREKSILKST IINDVRATVQE GKRYTNIGPS
 401 IVYRIRDAG QAVYEKYNNMFL PVHQEDYFW ITGTRSGLQQ CYRWLRIPLD
 451 KQLKADTFMG LREFLKGDEG RKRIVADATIK GAPAEIREOF MIAAENTNINI
 501 FAQKGYLGLD EFITSNLPKE QQQKMQGYFY EMLYGVNNAA LDETIIRRGL
 551 PEWQDQEARD RFPLIHSMDAY TGITLEYFAPM LLQLDGFSFV RSSGLQMTRS
 601 PGALIYLGLS VLVILGVITM FYVREKRAWW LFSDGKIRFA MSSARSERDL
 651 QKEFPKHVES LQRIGKDIH D*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

- 15 ORF88 shows 95.7% identity over a 371aa overlap with an ORF (ORF88a) from strain A of *N. meningitidis*:

			10	20	30
	orf88.pep		MVFLNADNGILVQDLPFEVKLKKFHIDFYN		
20	orf88a	AKDFKPESILGASNLSFRGNVNISEQGOSADVVF			
		210 220 230 240 250 260			
25	orf88.pep	TGMRDFASDIEVTDKATGEKLERTIRVNHP	40 50 60 70 80 90		
	orf88a	TGMRDFASDIEVTDKATGEKLERTIRVNHP	270 280 290 300 310 320		
30	orf88.pep	ASREPVVLKATSIHQFPLEIGKHKYRLEFDQFT	100 110 120 130 140 150		
	orf88a	ASREPVVLKATSIHQFPLEIGKHKYRLEFDQFT	330 340 350 360 370 380		
35	orf88.pep	TQEGRHYTINXXXXXXYRIRDAFGQAVEYKNYMLPV	160 170 180 190 200 210		
	orf88a	TQEGRHYTINXXXXXXYRIRDAFGQAVEYKNYMLPV	390 400 410 420 430 440		
40	orf88.pep	PLDKQLKADTFMALERFLKGDEGRKRVADATKGPA	220 230 240 250 260 270		
	orf88a	PLDKQLKADTFMALERFLKGDEGRKRVADATKGPA	450 460 470 480 490 500		
45	orf88.pep	DEQMLAAENTLNIFAQKGYL			
	orf88a	DEQMLAAENTLNIFAQKGYL			
50	orf88.pep	GLDEFITSNIKEQQDMQGYFYEMLYGVNNAALDET	280 290 300 310 320 330		
	orf88a	GLDEFITSNIKEQQDMQGYFYEMLYGVNNAALDET	510 520 530 540 550 560		
55	orf88.pep	TRYGLPFWQDEARNRFLHHSM	340 350 360 370		
	orf88a	TRYGLPFWQDEARNRFLHHSM	570 580 590 600 610 620		
60	orf88a	DAYTGLTEYFAPMLLQLDGFSEVRSSGLQMTRSXGP	ANVLFSDGKIRFAMMSSARSDLQKEFFKHVESLQLRGKDINHDX		
		LLVYVLGSVLLVLGTVLMFYVREKR	630 640 650 660 670		

The complete length ORF88a nucleotide sequence <SEQ ID 331> is:

```

 1 ATGAGTAAAT CCCGTGAGTC TCCCCCACIT CTTTCCCGTC CGTGGTTCGC
 51 TTTTTTCAGC TCCAAAGCCT TTGGGGTCGC TTTGCTGAGT CTGCTGGGTA
 101 TTGCACTGGT TATCGGTACG GTGTTGCGAC AAAACCAAGCC GCAGACGGAT

```

151 TATTTGGTCA AATTCCGATC GTTTTGGGCG CAGATTTTG GTTTCTGGG
 201 ACTGTATGAC GTCTCATGCTT CGGCATGGTT TGGCTTATC ATGATGTTTT
 251 TGCTGGTTTC TACCACTTTC TGCTGTATC GCAATGTGCC GCGCTCTGG
 301 CGGGAAATGA AGTCTTTTG GGAAAAGGTT AAAGAAAAAT CTTGGCGCC
 5 351 GATGGCCCAT TCTTCGCTT TGGATGTAAT AATTGGCCCC GAGGTTGCCA
 401 AACGTTATCTT GGAAGTACRA GGTTTTCAGG GAAAACCAT TAACCGTGA
 451 GACGGGTGGT TTCTGATGTC CGCCAAAAAA GGCACAAATGA ACAATGGGG
 501 CTATATCTT GCGCATGTTG TTCTGATGTT CATTGCTG TGCGGATGTTGA
 551 TAGACAGTAA CCTGCTGTT AAACCTGGT TGCTGACCGG TCGGATGTT
 601 CCGGACAATC AGGGCGTTTA TGCCAAAGGAT TTCAAGCCGG AAAGTATTTT
 651 GGGTGGTCC AATCTCTCAT TTAGGGCCTA CGTCAATATT TCCGAGGGG
 701 AGAGTGGCGA TTGTTGGTTTC TGCAATGCGG ACAACGGGGAT ATTGGTTCAG
 751 GACTTGCCTT TTGAACTGCA ACTGAAAAAA TTCCATATCG ATTITIACAA
 801 TACGGATGATG CGGGCGGATT TTGCGCATGTA TATTGAAAGTA ACGGATAAAGG
 851 CAACCGTGA GAAACTGGC CGCACCATCG CGGTGAAACCA TCCTTGCAC
 901 TTGCAAGCGCA TCACGGATTTC TGAGGGGAT TTGCGGAGG CGCGTTTCCGA
 951 TTGACATTC AAGGCCTGGG ATTGGGTGA TGCTTCGCGC GAGCCTGTCG
 1001 TTGTAAGGGC AAATCATCATA CACAGTTTC CGTGGAAAAT TGCGAACAC
 1051 AAATATCGTC TTGAGATCGA TGAGTTTACT TCTATGAAATG TGGAGGACAT
 20 1101 GCGGACGGGC GCGGAAGGG AAAAAGGCTT GAAATCCACG CTGAAACGTG
 1151 TCGGCCGCTG TACTCAGGAA GGTTAAAAAAAT ACACCAATAT CGGCCCTTC
 1201 ATGTTTACCA GTATCGTC GTCGGCGAAGG CGGGCGTGC ATATAAAAAA
 1251 CTATATGTCG CGGGTTTTC AGGAACAGGA TTATTTTTGG ATTACCGGCA
 1301 CGCGCAGGGG CTTCGACGAC CAATACCGT GCCTGCCTAT CCCCTGGAC
 1351 AAGCGTTGA AGGGCGACAG CTTTATGGCA TGCGGTGAGT TTGAAAGA
 1401 TGGGAAAGGG CGCAACAGTC TTGTTGCGGA CGCAACAAA GGGCACCTG
 1451 CGGAAATCCG CGGAAATTC ATGCTGCGTG CGGAAACAC GCTGAAACATC
 1501 TTGGCAACAAA AAGGCTTAT GGGATGCGAC GAATTATTA CGCTCAATAT
 1551 CGCGAAAGGG CAGCGAGATA AGATGCGAGG TTATTTCTAC GAAATGCTT
 1601 ACCGGCTGAT GAAACGCTG TTGGATGAA CCATACGCCG GTACGCCCTG
 1651 CCCGAATGGC AGCAGGTGA AGCGGGAGAT CGTTTCCCTGC TGCAACAGTAT
 1701 GGATGGTAC ACGGGGTTGA CGGAATATCC CGCGCCTATG TGCTGCAC
 1751 TTGATGGTTT TTCCGGAGTG GTCTGCTGG TTGTCAGAT GACCGCTTCC
 1801 CGCGCTGGC TTGGTGTCTA TCTCGGCTCG GTGCTGTGG TATTGGTAC
 1851 GTATTTGTTT TTATATGTCG CGGAAAACG CGCGTGGGTA TTGTTTTCAG
 1901 ACGGCAAAAT CGCTTGTGCG ATGCTCTGG CGCGCAGCGA ACGGGATTTG
 1951 CAGAGGAAT TTCCAARACCA CGTGGAGAGT CTGCAACGGC TCGGCAAGGA
 2001 CTTGAATCAT GACTGA

This encodes a protein having amino acid sequence <SEQ ID 332>:

40 1 MSKSRSSPPL LSRPWFAFFS SMRFAVALLS LLGIASVIGT VLQQNQPTD
 51 YLVKGSSFWA OIFGFGLDY VYASAWFVII MMELVSTSL CLIRNVPPFW
 101 REMKSREKTY KEKLAAAMRH SSSLLOVKIAA EVAKRYLEVO GFOGKINTIRE
 151 DGSVLLAAAKR GTMNWKYNGV AHVALIVICL GGLIDSNLLL KLGMLTRIV
 201 PDNQAYAKO FKPEISILGAS NLSFRGNVNII SEGOSADVVF LNADNGILVQ
 251 DLPEVFKLKD FHIDFYNTGM PROFASDIEV TKATGEKLE RTIRVNHPLT
 301 LHCINTLYQAS FADGGSDLTF KAWNLDGASR BPVVLKATSI HOFPLEIGKH
 351 KYRLERFDQFT SMNVEDMSEG AEREKSLKST LDVRRAVTOE GKYYTNICGPS
 401 IVYRDRADAG QAEVYKYNWY PFLQEDDFW IISGTRSGLQ QYRWLRIPLD
 451 RQLKADTMM LREFPKDQEG KRLVALDATK GRAPAIKEOF MLLAENTLNI
 501 FAQGYLGLD EFTINSNPK QDKMOMGYFY EMLYGVMNAA LDETRIRYGL
 551 PEQWQDEARD RFITTSNPKD LLGTEYPAFM LLQLDGESEV RSSEGLQMTRS
 601 PGALIVYLGS VLLVLTGTVLM FYVREKRAWV LFSDGDKIRFA MSSARSERDL
 651 QKEFPKHVES LQRIGKDNH D*

ORF88a and ORF88-1 100.0% identity in 671 aa overlap:

55 orf88a.pep MSKSRSSPPLLSRPWFAFFSSMRFAVALLSLLGIASVIGTVLQQNQPTDYLWKGSEFWA 60
 orf88-1 MSKSRSSPPLLSRPWFAFFSSMRFAVALLSLLGIASVIGTVLQQNQPTDYLWKGSEFWA 60
 60 orf88a.pep QIFGFLGLDYVYASAWFVIMMFVLVSTSLCLIRNVPPFWREMKSFREKVKEKSLAAAMRH 120
 orf88-1 QIFGFLGLDYVYASAWFVIMMFVLVSTSLCLIRNVPPFWREMKSFREKVKEKSLAAAMRH 120
 65 orf88a.pep SSSLLOVKIAPEVAKRYLEVQGFGQKTIINREDGSVLLIAAKGTMNKWGYIFAHVALIVICL 180
 orf88-1 SSSLLOVKIAPEVAKRYLEVQGFGQKTIINREDGSVLLIAAKGTMNKWGYIFAHVALIVICL 180
 orf88a.pep GGLIDSNLLKLGLMLTGRIVPDNQAVYAKDFKPESILGASNLSFRGNVNISEGQSADVVF 240

	orf88-1	G L D S N L L K L G M T G R I V P D N Q A Y A K D F K P E S I L G A S N L S F R G N V N I S E G Q S A D V F	240
5	orf88a.pep	LNA D N G I L V Q D L P F E V K L K F H I D F Y N T G M P R D F A S D I E V T D K A T G E K L E R T I R V N H P T	300
	orf88-1	L N A D N G I L V Q D L P F E V K L K F H I D F Y N T G M P R D F A S D I E V T D K A T G E K L E R T I R V N H P T	300
10	orf88a.pep	LHG I T I Y Q A S F A D G G S D L T F K A W N L G D A S R P V V L K A T S I H Q F P L E I G K H K Y R L E F D Q F	360
	orf88-1	L H G I T I Y Q A S F A D G G S D L T F K A W N L G D A S R P V V L K A T S I H Q F P L E I G K H K Y R L E F D Q F	360
15	orf88a.pep	S M N V E D M S E G A R E K S L K S T L N D V R A T Q E G K Y T N I G S I V Y I R D A A G Q A V E Y K N Y L	420
	orf88-1	S M N V E D M S E G A R E K S L K S T L N D V R A T Q E G K Y T N I G S I V Y I R D A A G Q A V E Y K N Y L	420
20	orf88a.pep	P V L Q E Q D Y F W I T G T R S G L Q Q Y R W L R I P L D K Q L K A D T F M A R E F L K D G E G R K R L V A D A T K	480
	orf88-1	P V L Q E Q D Y F W I T G T R S G L Q Q Y R W L R I P L D K Q L K A D T F M A R E F L K D G E G R K R L V A D A T K	480
25	orf88a.pep	G A P A E I R E Q P Q M L A A E N T L N I F A Q K G Y L G L D E F I T S N I P K E Q Q D K M Q Y Y E M L Y G V M A A	540
	orf88-1	G A P A E I R E Q P Q M L A A E N T L N I F A Q K G Y L G L D E F I T S N I P K E Q Q D K M Q Y Y E M L Y G V M A A	540
30	orf88a.pep	L D E T I R Y G L P E W Q D E A R N R F L L H S M D A Y T G L T E Y P A P M L L Q L D G F S E V R S G L Q M T R S	600
	orf88-1	L D E T I R Y G L P E W Q D E A R N R F L L H S M D A Y T G L T E Y P A P M L L Q L D G F S E V R S G L Q M T R S	600
35	orf88a.pep	P G A L L V Y L G S V L L V L G T V L M F V V E K R A W V L F S I G K I R F A M S S A R S E R D L Q K E F K H V E S	660
	orf88-1	P G A L L V Y L G S V L L V L G T V L M F V V E K R A W V L F S I G K I R F A M S S A R S E R D L Q K E F K H V E S	660
	orf88a.pep	L Q R L G K D L N H D	672
	orf88-1	L Q R L G K D L N H D	672

Homology with a predicted ORF from *N.gonorrhoeae*

ORF88 shows 93.8% identity over a 371aa overlap with a predicted ORF (ORF88.ng) from *N. gonorrhoeae*:

40	orf88.pep	M V F L N A D N G I L V Q D L P F E V K L K F H I D F Y N T G M P R D F A S D I E V T D K A T G E K L E R T I R V N H	60
	orf88ng	M V F L N A D N G I L V Q D L P F E V K L K F H I D F Y N T G M P R D F A S D I E V T D K A T G E K L E R T I R V N H	60
45	orf88.pep	P L T L H G I T I Y Q A S F A D G S D L T F K A W N L G D A S R P V V L K A T S I H Q F P L E I G K H K Y R L E F D	120
	orf88ng	P L T L H G I T I Y Q A S F A D G S D L T F K A W N L G D A S R P V V L K A T S I H Q F P L E I G K H K Y R L E F D	120
50	orf88.pep	Q F T S M N V E D M S E G A R E K S L K S T L P D Q R V A T Q E G H K Y T N X X X X X X Y I R D A G Q A V E Y K N	180
	orf88ng	Q F T S M N V E D M S E G A R E K S L K S T L P D Q R V A T Q E G H K Y T N X X X X X X Y I R D A G Q A V E Y K N	180
55	orf88.pep	Y M L P V L Q E Q D Y F W I T G T R S L Q Q Y R W L R I P L D Q L K A D T F M A R E F L K D G E G R K R V A D	240
	orf88ng	Y M L P V L Q E Q D Y F W I T G T R S L Q Q Y R W L R I P L D Q L K A D T F M A R E F L K D G E G R K R V A D	240
60	orf88.pep	A T K G A P A E I R E Q P Q M L A A E N T L N I F A Q K G Y L G L D E F I T S N I P K E Q Q D K M Q Y Y E M L Y G V	300
	orf88ng	A T K G A P A E I R E Q P Q M L A A E N T L N I F A Q K G Y L G L D E F I T S N I P K E Q Q D K M Q Y Y E M L Y G V	300
65	orf88.pep	N A A L D E T X T R Y G L P E W Q D E A R N R F L L H S M D A Y T G L T E Y P A P M L L Q L D G F S E V R S G L Q M T R S	360
	orf88ng	N A A L D E T I R Y G L P E W Q D E A R N R F L L H S M D A Y T G L T E Y P A P M L L Q L D G F S E V R S G L Q M T R S	360
	orf88.pep	T R S X G P L L V Y L	371
	orf88ng	T R S X G P L L V Y L	420

An ORF88ng nucleotide sequence <SEQ ID 333> was predicted to encode a protein having amino acid sequence <SEQ ID 334>:

5	1 MVFIAADNGM LVQDLPFEVK LKKFHIDFVN TGMPPRDFASD IEVTDKATGE 51 KLERITIRVNH PLTLMNTGTT QASFPFDGGSD LTIFGAWNLRD ASREPVVLKA 101 TSIHQFPLEI GKHKYRLIED QFTSMVNEDM SEGAEREKSL KSTIANDVRAS 151 TQEKGKYYNTI GPSIVYRIRD AAGQAVEYKG YMLFILQDKD YFWLTTGRSG 201 LQQQRWLRIT PLDQLKLADT EMALEFLFKD GEGRKRRLVAD ATKDAPABIR 251 EQFMLAABNTI LINIFAQKGYL GLDEFITSNPK FGKQODKMCG YFYEMLYGV 301 NAALDETIRY YGLPEWQDQE ARNRFLHNS DAYTGLTEYP APMLLQLQDG 351 SEVRSSRLQM TRSPGALIVY LGSVLVLVLT VEMFYVVKRR ANWLFSNKKI 401 RFAMSSARSTI RDLQKEFFPKH VESLQLRKD LNHD*
---	--

Further work revealed the complete gonococcal DNA sequence <SEQ ID 335>:

15	1 ATGAGTAATAT CCCGTATAT TCCACACATT CTTCCCGTC CGTGGTTCGC 51 TTTTTTCAGC CTATCGGCT TTGCGGTGCG TTGCTCATG CTGCTGGGTA 101 TTGATCGCGT ATTTCGCAAC GTCATCACAG AAAACAGCC GCAGACGGAT 151 TATTGTCGA AATTTCGCAAC GTTTCGACT CGGATTTTG ATTTTTGGG 201 TTGTTATGAT GTCATGATCT CGGCATGGTT TGTCGTTATC ATGATGTTTC 251 TGGTGGTTTC TACCGCTTGC TTGTTAACTC GTAACTGTTCC GCGCTTTGG 301 CGCGAAATGA AGCTTCTTCG GAAAAGGGT AAAGAAAAAT CTCGGCGC 351 GATGGCCCAT TCTTCGCTCT TGGATGTTAA ATTTCGGCCCC QAAGTTGCCA 401 AACGGTTATCTC GGAGCTGCGG TTGTTTCAGG GAAAACCCG CACCGCTGAG 451 GACGGTTGCG TCTCTGATCG CGCCAAAAGG GGCACaaatga acaaATGGGG 501 CTATATCTTTC CGGcaatgtg ctTTGATGTT CATTTCGCTG GCGGGTTGGA 551 TAGACGAACT CCTCTGCTCTG AACCTGGCTG TGCTGGATG TTGGGATTTG 601 CGCGAACATC AGGGCGGTG TTGCAAGATA TTCAAGCGCG AAAGTATTTT 651 GGGTGGCTTAA ATCTTCCTAT TTAGGGCAAA CTGCAATATT TCGCAGGGGC 701 AAAGTCGGGA TTGTTGGTTCT CTGATGCGC ACAACGGGT GTTGGTTCA 751 GACTTGCCTT TTGAGAGCTA ACTGAAAAAAT TTCCATATTCG ATTTTTACAA 801 TACGGGTATG CGCGCGCAT TTGCGAGGA TTGAGAAGTA ACCGACAAGG 851 CAACGGCTGA GAAACTGAG CGACCATCC CGGTGAAACCA TCTTTTGACC 901 TTGCGACCGCA TCRGCGATTA TTGCGAGATG TTGGCGACG GCGGTTCCGA 951 TTGACATTC AGGGCGTGGG ATTGGGGGA TGCTTCGGCC GACACTGTCG 1001 TTGTGGAGGC AACCTCCATC CACCGTTCTC TTGAGGAAAT CGCAAACAC 1051 AAATATCGTC TTGAGTCCTC TGATGCTC TGATGAAATG TTGAGGAGAT 1101 GAGCGAGGGT GGGGAGCGGG AAAAAGGCCT GAATTCACG CTGACGATG 1151 TCCGGCGGTG TACTCAGGA GGTTAAAAAAT ACACCAATAT GGGCCCTTCC 1201 ATCGTGTACG CGATCTGCGA TGcgcAGGG CAGGGCGTGC AATATAAAA 1251 CTATATGTC CGGATTTCG AGGACAAGAA TTATTTTGG CTGACGGCCA 1301 CGCGACAGGG CCTTCACAGC CAATACCGCT CGCTCGTAT CCCCCCTGGAC 1351 AACGACTGGA AGCGCGCAC CTTATGCGA TTGGCTGAGT TTTGAAAGA 1401 TGGGGAAGGG CCCAACAGTC TTGTTGGCGG CGCACACAAA GAGGCACCTG 1451 CGGAATTCGG CGAACATTC ATGCTGCTG CGGAAACAC GCTGAATATC 1501 TTGGCGCAAAG AACGGCTATTG GGGATTGGAC GAATTATTTA CGTCAAAATAT 1551 CGCGAAAGGG CAGCAGGATA AGATGCAAGG CTATTTCTAC GAATGCTTT 1601 ACGGGCTGTG GRAGCGCTGT TTGGATGAAAG CCATACCGCCG GTACGGCTTG 1651 CGCGATGCGC AGCGCGATG AGCGCGGACG CTTTCTGC TGACAGATAT 1701 GGGATGCTAT CGGGGGCTGA CGGAATATCC CGCGCCCTATG CTGCTCCAGC 1751 TTGACGGGGTT TTGGCAGGTT CGTCTCTAG TTGGCAGAT GACCCCTTCG 1801 CCGGGTGGCGC TTTTGTGCTA Tctcgctcg gtatggttg TTTGGgtac 1851 ggatTTtagt tTTTATGTGC CGGAAAACCG GGCGCTGGgtat TTGTTTCTag 1901 aCGGCAAAAT CGGTGTTGCT ATGCTCTGgg cccqcaacgca ACGGGATTTG 1951 cAGGaaaggat TTCCAANACCA CGtcgAGAGC CTGCAACGgc tcggcaaggA 2001 CttggaaATCAT GACTga
----	--

This corresponds to the amino acid sequence <SEQ ID 336; ORF88ng-1>:

55	1 MSKSRISPLT LSRPWFAFFS SMRFAVALLS LLGIASVIGT VLQONOPQTD 51 YLVFKGPFTT RIEFDLGLND VYASAWFVII MMELVSTSL CLIRNVPPFW 101 REMKSFRKEV KEKSLSAAMHR SSSLDVVKIAP EVAKRYLEV RFOGKTVSRE 151 DGSVLLIAAKK GTMNWKWYIE AQVALVICL GGLIDSNILL KLGMLAGRIV 201 FENQVYAKD FKRPSILGALS NLSEGRGNVNI SEGOSADWVF LNADNGMLVQ 251 DLPEVFKLKK FHIDFVNTGM PRDFASDIEV TIKATGEKLE RTIRVMHPLT 301 LHGTTIYQAS FADGGSDFLT KANNRLDASR EPVVLKATSI HOFPLIEIGKH 351 KYRLEFDQFT SMNVEDMSSEG AEREKSLKST LNDVRAVTQE GKKYTNIQPS 401 IVYVIRDAAG QAEVYKNYML PIQOKDWFW LTGTRGSLQO QYWRLNLIPLD 451 KOLKADTFTM LREFLKGEG KKLRLVADATK DAPABIREQF MLLAENTINI 501 FAQKGYLGLD EFITSNIPKG QDKRMQGYFY EMLYGVNMAA LDETiRRYGL
----	--

```

551 PEWQDEARN RFLLHSMDAY TGLTTEPAPM LLQLDGFSEV RSSGLQMTRS
601 PGALLVYLG S VLLVLGTVEM FYVREKRAWV LFSDGKIRFA MSSARSERDL
651 QEKFPKRVES LQRIGKDLNH D*

```

ORF88ng-1 and ORF88-1 show 97.0% identity in 671 aa overlap:

5	orf88-1.pep	MSKSRRSPPLLSRPWF A FFSM RFAV ALLS LLLG IAS VIG TVL QL QQNP QT DYL VKFG SEWA 60
	orf88ng-1	MSKSRRISPT LLSRPWF A FFSM RFAV ALLS LLLG IAS VIG TVL QL QQNP QT DYL VKFG SEWT 60
10	orf88-1.pep	QTFGLFLGYD YVA SAW FV INMFLV VST SCL IRN VPFP WREM KS F RE KV KE KSLA MMRH 120
	orf88ng-1	RTFDFGLFLGYD YVA SAW FV INMFLV VST SCL IRN VPFP WREM KS F RE KV KE KSLA MMRH 120
15	orf88-1.pep	S SLL DV KIA P EVA K RY LEV Q G F Q G K T I N R E D G S V L I A A K G T M M N K W G Y I F A H A V I V I C L 180
	orf88ng-1	S SLL DV KIA P EVA K RY LEV Q G F Q G K T I N R E D G S V L I A A K G T M M N K W G Y I F A Q V A L I V I C L 180
20	orf88-1.pep	G GLI D S N L L L K L G M L T G R I V P D N Q A V Y A R D Q F R P E S I L G A S N I S F R G N V N I S E G Q S A D V F F 240
	orf88ng-1	G GLI D S N L L L K L G M L A G R I V P D N Q A V Y A R D Q F R P E S I L G A S N I S F R G N V N I S E G Q S A D V F F 240
25	orf88-1.pep	L N A D O N G L I V Q Q D L P F E V K L K K F H I D F Y N T G M P R D F A S D I E V T D K A T G E K L E R T I R V N H P L T 300
	orf88ng-1	L N A D O N G L I V Q Q D L P F E V K L K K F H I D F Y N T G M P R D F A S D I E V T D K A T G E K L E R T I R V N H P L T 300
30	orf88-1.pep	L H G I T T Y Q A S F A D G G S D L T F K A W N L G D A S R E P V V L K A T S I H Q F P L E I G K H K Y R L E F D Q F T 360
	orf88ng-1	L H G I T T Y Q A S F A D G G S D L T F K A W N L G D A S R E P V V L K A T S I H Q F P L E I G K H K Y R L E F D Q F T 360
35	orf88-1.pep	S M N V E D M S E G A R E K S L K S T L N D V R A T V T Q E G K Y T N I G P S I V Y R I D A A G Q A V E Y K N Y M L 420
	orf88ng-1	S M N V E D M S E G A R E K S L K S T L N D V R A T V T Q E G K Y T N I G P S I V Y R I D A A G Q A V E Y K N Y M L 420
40	orf88-1.pep	P V L Q F Q D Y F W I T G T R S G L Q Q Q Y R W L R I P D L K Q L K A D T F M A L R E F L K D G E G R K R L V D A T K 480
	orf88ng-1	P V L Q F Q D Y F W I T G T R S G L Q Q Q Y R W L R I P D L K Q L K A D T F M A L R E F L K D G E G R K R L V D A T K 480
45	orf88-1.pep	G A P A E I R E Q F M L A A E N T L N I F A Q K G Y L G L D E F I T S N I P K E Q Q D K M Q G Y F Y E M L Y G V M N A A 540
	orf88ng-1	D A P A E I R E Q F M L A A E N T L N I F A Q K G Y L G L D E F I T S N I P K G Q D D R M Q G Y F Y E M L Y G V M N A A 540
50	orf88-1.pep	L D E T I R R Y G L P E W Q Q D E A R N R F L L H S M D A Y T G L T E Y P A P M L L Q L D G S E V R S S G L Q M T R S 600
	orf88ng-1	L D E T I R R Y G L P E W Q Q D E A R N R F L L H S M D A Y T G L T E Y P A P M L L Q L D G S E V R S S G L Q M T R S 600
55	orf88-1.pep	P G A L L V Y L G S V L V L G T V L M F Y V R E K R A A V L F S D G K I R F A M S S A R S E R D L Q K E F P K H V E S 660
	orf88ng-1	P G A L L V Y L G S V L V L G T V F M Y V R E K R A A V L F S D G K I R F A M S S A R S E R D L Q K E F P K H V E S 660
60	orf88-1.pep	L Q R L G K D L N H D 671
	orf88ng-1	L Q R L G K D L N H D 671

Furthermore, ORG88ng-1 shows homology with a hypothetical protein from *Aquifex aeolicus*:

55	gi 2984296 (AE000771) hypothetical protein [Aquifex aeolicus] Length = 537
	Score = 94.4 bits (231), Expect = 2e-18
	Identities = 91/334 (27%), Positives = 159/334 (47%), Gaps = 59/334 (17%)
60	Query: 16 FAF F S M R F A V A L L S L L G I A S V I G - T V L Q L Q N Q P Q T D Y L V K F G P W T R I F D F L G L Y D V Y A S 74
	+ F + S++ A+ ++ + LGI S++G T ++ Q N Q Y L + F G L D V + S
	Sbjct: 80 Y D P L S L K L A I F I M V L G I L S M L G S T Y I K Q N Q S F E W Y L D Q F G Y D V G I W T K L W L N D V F H S 139
65	Query: 75 A W F V I M M F L V U S T S C L I R N V P P W R E M K S F R E K V K E K S L A A M R H S S L L D V K I A P E V A K 134
	+ + + + + L V + + W + + S + E + + + A + H + + V K I P + K
	Sbjct: 140 W Y Y I F V L I L A V N I F C S T K I R L P R V W Q A F S - K E R I L K D E H A E K U L K F I T V K I - P D K K I 197
70	Query: 135 -- R Y L E V R G F Q Q G K T V S R E D G S V L I A A K G T M M N K W G Y I F A Q V A L I V I C L G G L I D S N L L K L K 192
	+ L + G F + V E + + A + K G + + G + A L + V I G L D
	Sbjct: 198 V L K F L L K K G F K V F V E E G N K L Y V F A E K G R F S R L G V Y I T H I A L L V I M A G A L I D ----- 249

Query: 193 GMLAGRIVPDNQAVYAKDFKPESIYLGSNLFRGNVNISEGQSADVVFLNADNGLMVQLD 252
+1+G RG++ ++EG + DV+ + A+ L
-----AIVGV-----RGLSLIVAEQGDTNDVMLVGAE--QKPYKL 280

5 Sbjct: 250 -----
Query: 253 PFEVKLKEHIDFY---NTGMPRDEA-----SDIEVTOKATGEKLER--TIRVNHPLT 300
PF V L F I Y N + + FA SDIE+ + G K+E T+VN P
Sbjct: 281 PFAVHLIDFRIKYABENNVYDKRFAQAVSSYESDIEIIN--GGKVEAKGTVKVNEPFD 337

10 Query: 301 LHGIFTIQASFA--DGGSDLTFKAQNLRDAPR 332
++QA++ DG S + + A +P
Sbjct: 338 FGRYRLFQATYGIQDGTSGMGIVIVDRKKAHEPD 371

Based on this analysis, including the putative transmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could 15 be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 40

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 337>:

20 1 ATGATGAGTA ATAmAATGGG ACAAAAAGGG TTTACATTGA TTGmGmTGAT
51 GATAGTCGTC CGCATACTCG GCATTTATCAG CGTCATGCC ATACCTCTT
101 ATCmAGTTA TATIGAAAAAA GGCTATCAGT CCCAGCTTTA TACGGAGATG
151 GYCCTATCA ACATAATTTC CAACAGCTTT ATTTCGAAAA ATCCCTGG
201 CGATATTCAG ACCATCGAGA CAAACACTGGA ATATTTTGTG TCAGGCCTATA
251 AGATGAATCC AAAATTGCCC AAAAATATA GTGTTTCCGT AAAGTTTGTG
301 GATAAAGAAA PATCAAGGGC ATACAGTTG TTGCGGGTTG CGAACGGGG
351 GACGGGTTAT ACTTTGTCGG TATGGATGAA CAGGGTGGGC GACGGATACA
401 AATGCCGTTGA TGCGCTTCT GCCCAAGGCC ATTGGAGAC CTTGCTCTCA
451 GATGTCGGCT GTGAAGCCTT CTCTAATCGT AAAAATATA

This corresponds to the amino acid sequence <SEQ ID 338; ORF89>:

30 1 MMSNNMXQKG FTЛИXXMIVV AILGIISVIA IPSYXSYIEK GVQSQLYTEM
51 XGINNNSKQF TLKNPLDQNQ TIENKLSIFV SGYMMNPKIA KKYSVVKFV
101 DEKEKSRAYRL VGVPKAGTGY TLSVWMNSVG DGYKCRDAAS AQAHLETLSS
151 DVGCEAFSNR KK*

Further work revealed the complete nucleotide sequence <SEQ ID 339>:

35 1 ATGATGAGTA ATAAARTGGG ACAAAAAGGG TTTACATTGA TTGAGATGAT
51 GATAGTCGTC CGCATACTCG GCATTTATCAG CGTCATGCC ATACCTCTT
101 ATCmAACTTA TATIGAAAAAA GGCTATCAGT CCCAGCTTTA TACGGAGATG
151 GTCGGTATCA ACATAATTTC CAACAGCTTT ATTTCGAAAA ATCCCTGG
201 CGATATTCAG ACCATCGAGA CAAACACTGGA ATATTTTGTG TCAGGCCTATA
251 AGATGAATCC AAAATTGCCC AAAAATATA GTGTTTCCGT AAAGTTTGTG
301 GATAAAGAAA PATCAAGGGC ATACAGTTG TTGCGGGTTG CGAACGGGG
351 GACGGGTTAT ACTTTGTCGG TATGGATGAA CAGGGTGGGC GACGGATACA
401 AATGCCGTTGA TGCGCTTCT GCCCAAGGCC ATTGGAGAC CTTGCTCTCA
451 GATGTCGGCT GTGAAGCCTT CTCTAATCGT AAAAATATA

45 This corresponds to the amino acid sequence <SEQ ID 340; ORF89-1>:

1 MMSNNMXQKG FTЛИEMMIVV AILGIISVIA IPSYQSYIEK GVQSQLYTEM
51 XGINNNSKQF TLKNPLDQNQ TIENKLSIFV SGYMMNPKIA KKYSVVKFV
101 DEKEKSRAYRL VGVPKAGTGY TLSVWMNSVG DGYKCRDAAS AQAHLETLSS
151 DVGCEAFSNR KK*

50 50 Computer analysis of this amino acid sequence gave the following results:

Homology with PilE of *N. gonorrhoeae* (accession number Z69260).

ORF89 and PilE protein show 30% aa identity in 120a overlap:

-226-

orf89	8	QKGKFTLIXMMIVVAILGGSVIAIAPSYSYIEKGYQSQLYITEMGXINNISKQFILKNPL- QKGKFTL MIV+AI+GI++A+P+Y Y + S+ G + ++ L + +	66
PILE	5	QKGKFTLIELMIVIAIVGILAVALPAFYDITARAIVQSVSEALLAEGQKSAVTEYHNGW I	64
orf89	67	-DDDNQTENKEIFVSGYKMNPKIAKKYVSFKVFDREKSRAYRLGVGPAGTYGTLSW DN + + + G + KI KY SV + + GV G LSW+K	125
PILE	65	KPDNTS-----AVGASSKKIGKYGKVSVQKVGKTEAMASTGVNKIEQGKLNLL I	115

Homology with a predicted ORF from *N.meningitidis* (strain A)

- 10 ORF89 shows 83.3% identity over a 162aa overlap with an ORF (ORF89a) from strain A of *N. meningitidis*:

			10	20	30	40	50	60
15	orf89.pep	MMSNNXKQKGTLIXXMMIVAILGLIISVIAIPSYKSYIEKGYSQSLTYTEMXGNNNISKQQF						
	orf89a	MMSNKMEQRKGFTLIXXXXAIXXXXSVXXXXYSYIEKGYSQSLTYTEMVGNNNISKQW	10	20	30	40	50	60
20	orf89.pep	ILKNPLDDNQTINKELFVSGYKMNPKIAKYSVSVKFVDEKSRAYRLVGVPKAGTGY	70	80	90	100	110	120
	orf89a	ILKNPLDDNQTIKSKLELFVSGYKMNPKIAEYKNVSVHFVNKEPRAYSLVGVPKTGTGY	70	80	90	100	110	120
25	orf89.pep	TLSVWMNSVGDGYKCRDAASQAHLLETLTSSDVGCEAFSNRKXX	130	140	150	160		
	orf89a	TLSVWMNSVGDGYKCRDAASARAHLETLTSSDVGCEAFSNRKXX	130	140	150	160		

The complete length ORF89a nucleotide sequence <SEQ ID 341> is:

30	1	ATGATGAGTA	ATAAAATGGA	ACAAAAAAGGG	TTTACATTGA	TTGNGANGNT
	51	NATNCNCCTC	GGCATACNCN	GNCNTTACCG	CSTCATTCN	ANNNTCCTTC
	101	ATCNAGGTTA	TAGTTGAAA	GGCTATCACT	CCACGGCTT	TACGGAGATG
	151	GTCTGGTATCA	AACATATTAT	CAACAGCTNT	TTTTTGAAA	ATCCCTCCG
	201	CGATAATCG	ACCATCAAGA	CGAACATGGA	ATAATTTCGG	TCAGGGCTATAC
35	251	AGATGAATTC	GRAAAATTGCC	GAAAATATTA	ATTTGTTCTGT	GGATTTGTTG
	301	AATGAGGAA	AACCHNAGGG	ATACAGCTTS	GTGCGGGT	CAAAGACGGG
	351	GCAGGGTTAT	ACTTTGTCGG	TATGGGTA	CGGGGGCCC	GACGGGATACA
	401	ATTCGGCTGA	TGGCGCTT	GGCCGGGCC	ATTTGGAGAC	CTTGGCTCTA
	451	GATGTCGGCT	GTGAGGCTT	CTCTAATGCT	AAAAAATAG	

- 40 This encodes a protein having amino acid sequence <SEQ ID 342>.

1 MMSNKMEQKG FTLIXXXXXX AIXXXSVIX XXXYXYIEK GYQSQLYTEM
51 VGINNISKQX ILKNPLDDNQ TIKSKLIEFV SGYKMNPKIA EKYNVSVHFV
101 NEEKPRAYSL VGVPKTGTY TLSWVMNSVG DGYKCRDAAS ARAHLETLSS
151 DVGCEAFNSR KK*

45 ORF89a and ORF89-1 show 83.3% identity in 162 aa overlap:

		10	20	30	40	50	60
50	orf89a.pep	MMSNMKMEQKGFTLIXXXXAIXXXXXSVIXXXXXXSYIEKGYQSQLYTEMVGINNISKQXW					
	orf89-1	MMSNMKMEQKGFTLIEMMIVVAILGIISVIAIPSYQSIEKGYQSQLYTEMVGINNISKQF	10	20	30	40	50
55	orf89a.pep	ILKNPLDDNOTIKSLETFVSGYKMNPKIAEKYNNVSHFWNEEKFRAYSLVGVPKTGTGTY	70	80	90	100	110
	orf89-1	ILKNPLDDNTIENKLETFVSGYKMNPKIAKYSVSKVFDKEKSRAYRLVGVPKAGTGY	70	80	90	100	110
60	orf89a.pep	TLSWVNNSWGDGYKGCRDAASARAHLETISSLSDVGCEAFSNRKKX	130	140	150	160	
	orf89-1	TLSWVNNSWGDGYKGCRDAASAOAHLETISSLSDVGCEAFSNRKKX					

130 140 150 160

Homology with a predicted ORF from *N.gonorrhoeae*

ORF89 shows 84.6% identity over a 162aa overlap with a predicted ORF (ORF89.ng) from *N.*

5 *gonorrhoeae*:

orf89	MMSNMXQKGFTLIXMIVVAILGIISVIAIPSYXSYIEKGYQSQLYTEMXGINNNISKQF	60
orf89ng	MMSNMXQKGFTLIEEMIVVAILGIISVIAIPSYXSYIEKGYQSQLYTEMXGINNNISKQF	60
10 orf89	ILKNPLDDNQTIEENKLIEFVSCYKMNPKIAKKYSVSVKFDKEKSRAYRLVGVPKAGTGY	120
orf89ng	ILKNPQDDNTLKSKLIEFVSCYKMNPKIAKKYSVSVRFVDAEKPRAYRLVGVPNACTGY	120
15 orf89	TLSWMNSVCGDGYKCRDAASAQAHLETLLSDVGCEAFSNRKK	162
orf89ng	TLSWMNSVCGDGYKCRDATSAQAYSDTLSADSGCEAFSNRKK	162

The complete length ORF89ng nucleotide sequence <SEQ ID 343> is:

1	aTGATGAGCA	ATAAAATGGA	ACAAAAAGGG	TTTACATTGA	TTCAGATGAT
51	GATAGCTTC	ACGATCATCG	GCATCATCAG	CCTCATTCGCC	ATACCTTCTT
101	ATCACAGGTA	TATTTGAAA	GCCATTCAGT	CCACAGCTTA	TACCGAGATG
151	GTGCGATCATCA	ACCATTTCTT	CAACAGTTC	ATTTTGAAAA	ATCCCCAGGA
201	CGTATATGAT	ACCCCTCAAGA	GCAACATGA	AATATTTCTC	TCAGGCTATA
251	AGATGAATG	GAAPATTGCC	AAAAAATATA	CTCTTTCGgt	aaggtttCTC
301	ga:CCGAAA	AACCAAGGGC	ATACAGCTTC	CTCGGGCTTC	CGAACCCGGG
351	GACGGGTAT	ACTTGTCTG	TATCGATCAA	CAGCCTGGCC	GRAGGATACA
401	AATGGGGTGA	TGCCACTCT	GGCCAGGCT	ATTCGGACAC	CTTGTCCCCA
451	GATAGGGCT	GTGAGGCTT	CTCTAATGCT	AAAAAATAG	

This encodes a protein having amino acid sequence <SEQ ID 344>:

1	MMSNMXQKG	FTLIEEMIVV	TILGIISVIA	IPSYXSYIEK	CYQSQLYTEM
51	VGINNLVK	ILKNPLDDN	TLSKSLKIFV	SCYKMNPKIA	KKYSVSVRFV
101	DAEKPRAYRL	VGVPNACTCY	TLSWMNSVNG	DCYKCRDATS	AQAYSDTLSA
151	DSCCEAFSNR	KK*			

This gonococcal protein has a putative leader peptide (underlined) and N-terminal methylation site (NMePhe or type-4 pili, double-underlined). In addition, ORF89ng and ORF89-1 show 88.3% identity in 162 aa overlap:

10	orf89-1.pep	MMSNMXQKGFTLIEEMIVVAILGIISVIAIPSYXSYIEKGYQSQLYTEMXGINNNISKQF	60
40	orf89ng	MMSNMXQKGFTLIEEMIVVAILGIISVIAIPSYXSYIEKGYQSQLYTEMXGINNNISKQF	60
70	orf89-1.pep	ILKNPLDDNQTIEENKLIEFVSCYKMNPKIAKKYSVSVKFDKEKSRAYRLVGVPKACTGY	120
45	orf89ng	ILKNPQDDNTLKSKLIEFVSCYKMNPKIAKKYSVSVRFVDAEKPRAYRLVGVPNACTGY	120
130	orf89-1.pep	TLSWMNSVCGDGYKCRDAASAQAHLETLLSDVGCEAFSNRKX	160
50	orf89ng	TLSWMNSVCGDGYKCRDATSAQAYSDTLSADSGCEAFSNRKX	160

Based on this analysis, including the gonococcal motifs and the homology with the known PilE protein, it was predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF89-1 (13.6kDa) was cloned in the pGex vector and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 11A shows the results of affinity purification of the GST-fusion protein. Purified GST-fusion protein was used to immunise mice, whose sera gave a positive result in the ELISA test., confirming that

5 ORF89-1 is a surface-exposed protein, and that it is a useful immunogen.

Example 41

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 345>:

10 1 ATGAAAAAAAT CCTCCCTCAT CAGCGCATTG GGCATCGGTA TTTTGGACAT
 51 CGGCATGGCA TTGCCCCCCC CTGCCGACGC GTTAAGCCAA ATCCGTCAA
 101 ACGCCACTCA AGTATTGAGC ATCTTAAAAAA ACGGCGATGC CAAACACCGCT
 151 CGCCAAAAGG CGGAAGGCATA TGCGGATTCC TATTTTCGATT TCCAACGTTAT
 201 GACCCGATTG GGGTTCGGCA ACCCTTGGGCG CACCGGTCAC GACG GCAA
 251 AACAGCGTT GGCGn...AGAA TTTCACCC...

This corresponds to the amino acid sequence <SEQ ID 346; ORF91>:

15 1 MKKSSLISAL GIGILSIGMA FAAPADAVSQ IRQNATQVLS ILKNGDANT
 51 ROKAEAYAIP YDFPQRMAL AVGNPWXTXS DXQKQALAKE FQP...

Further work revealed the complete nucleotide sequence <SEQ ID 347>:

20 1 ATGAAAAAAAT CCTCCCTCAT CAGCGCATTG GGCATCGGTA TTTTGGACAT
 51 CGGCATGGCA TTGCCCCCCC CTGCCGACGC GTTAAGCCAA ATCCGTCAA
 101 ACGCCACTCA AGTATTGAGC ATCTTAAAAAA ACGGCGATGC CAAACACCGCT
 151 CGCCAAAANG CGGAAGGCATA TGCGGATTCC TATTTTCGATT TCCAACGTTAT
 201 GACCCGATTG GGGTTCGGCA ACCCTTGGGCG CACCGGTCAC GACG GCAA
 251 AACAGCGTT GGCCAAAAGG TTTCACCCACCC TCGCTGATCCG CACCTATTCC
 301 GGCACGATGC TGAATTTAAAC AAACGGCACR GTCAACGTCA AGAACACATCC
 351 CATCAGCTCAT AAAGGGCGCA RAGAAATCAT CGTCGGGGCC GAAGTGGCCG
 401 TACCCGGCCA AAACCCGGTC AACATGACT GTACCCACTA CCAAGCGGC
 451 GGTAAATACC YTACCTTACA CGTCGGCATC GAAGGGCGCA GCTCTGGTTAC
 501 CGTGTACCGC ACCAAATTG GCGAAATTAT CAAAGCGAAA GGGCTGAGC
 551 GACTGATTGC CGAGTTGAAA GCCAAAAAACG CGGCCAAMATA A

30 This corresponds to the amino acid sequence <SEQ ID 348; ORF91-1>:

1 MKKSSLISAL GIGILSIGMA FAAPADAVSQ IRQNATQVLS ILKNGDANT
 51 ROKAEAYAIP YDFPQRMAL AVGNPWRTAS DQOKQALAKE FOTLLIRIYTS
 101 CTMLKLKNAN UNVKDNPIVN KGGKELIIVRA EVGVPGQKPV NMDDTTTYQSG
 151 GKTRTYNVAI EGASLTVVR NQFGSEIIKAK GUGLLIAEIK AKNGK*

35 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF91 shows 92.4% identity over a 92aa overlap with an ORF (ORF91a) from strain A of *N. meningitidis*:

40	orf91.pep	10 20 30 40 50 60
	orf91a	: : : : : : : : : : MKKSSLISALGIGILSIGMAFAAPADAVSQIRQNATQVLSILKNGDANTAROKAEAYAIP
45	orf91.pep	10 20 30 40 50 60
	orf91a	70 80 90 YDFPQRMALAVGNPWXTSDXQKQALAXEFQP : : : : : YDFPQRMALAVGNPWRTASDAQKQALAKEFQTLLIRIYTSGTMLKLKNANVNVKDNPIVN

orf91a KGGKEIIVRAEVGVPGQKPVNMDFTTYQSGGKYRTYNVAIEGASLTVYRNQFGEIIKAK
130 140 150 160 170 180

The complete length ORF91a nucleotide sequence <SEQ ID 349> is:

This encodes a protein having amino acid sequence <SEQ ID 350>.

1 MKKSSFISAL GIGILSIGMA FAAPADAVNQ IRQNATQVLs ILKSGDANTA
51 RQKAERYAAN YDFQRMAL AVGPNWRTA DAQKQALAKE FQFLIRITYS
101 GTMLKLKAYN VNVKONPIVN KGGKEIIVRA EVGVPGQKPV NMDFTTYQSQQG
151 GKVRVYVNAJ EGASLUTVYR NOFEGETIKAK GVDGLIALE AKANGSK-

QRF91a and QRF91-1 show 98.0% identity in 196 aa overlap:

Homology with a predicted ORF from *N.gonorrhoeae*

ORF91 shows 84.8% identity over a 92aa overlap with a predicted ORF (ORF91.ng) from *N. gonorrhoeae*.

The complete length ORF91ng nucleotide sequence <SEQ ID 351> is predicted to encode a protein having amino acid sequence <SEQ ID 352>:

-230-

```

1 VKKSSFISAL GIGILSIGMA FASPADAVGQ IRQNATQVLT ILKSGDAASA
51 RPKAAYAWP YFDQRMRTAL AVGNPWRATAS DAQKQALAKE FQTLLIRIYTYS
101 GTMLKFKNAT VNVKDNPIVN KGGKEITVURA EVGIPGOKPV NMDFTTYQSG
151 GKYRTYNVAI EGTSLVTVYR NQFGEIIKAK GIDGLIAELK AKNGKK*

```

5 Further work revealed the complete nucleotide sequence <SEQ ID 353>:

```

1 ATGAAAAAAAT CCTCCCTTCAT CAGCGCATTG GGCATCGGT A TTTTGGACAT
51 CGGCATGGCA TTTCGCTCCC CGGCCGACGC AGTGGACACA ATCCGCAAA
101 ACGCCAACACA GGTITTGACCC ATCCTCAAAA CGGGCCGACCC GGTTTCTGCA
151 CGCCCAAAAG CGGAGCTTA TGCGGTTCC TATTTCGAT TCCAAGCTAT
201 GACCCGATTG CGGTGTCGGA ACCCTTTGGG TACCGGCTCC GAGGGCCAAA
251 AACAGCCTT GCGGAAAGGA TTTCAAACCC TGCTGATCCC CNCTATCTC
301 GGCACGATGC TGAATTCAA AAACCGGACG GTCAACGTC AAGAACATCC
351 CATCTCAAT AGGGCGGCA AGGAAATCTG GTCTCGTGC GAAAGTCGCA
401 TCCCCGGCA GAAGGGCGTC AATAAGGACT TTACCAACTA CCAAGCGGC
451 GGGAAATACCG TAATCCCTA CGTGGCCATG AAAGGCGACG GCTCTGGTTAC
501 CGTGTACCGC AACAAATCTG GCGGAAATCAT CAAAGCCAAA GGCATCGACG
551 GGCATGATTGC CGACTTGAAGA GCGGAAATAA A

```

This corresponds to the amino acid sequence <SEQ ID 354; ORF91ng-1>:

```

20 1 MKKSSFISAL GIGILSIGMA FASPADAVGQ IRQNATQVLT ILKSGDAASA
51 RPKAAYAWP YFDQRMRTAL AVGNPWRATAS DAQKQALAKE FQTLLIRIYTYS
101 GTMLKFKNAT VNVKDNPIVN KGGKEITVURA EVGIPGOKPV NMDFTTYQSG
151 GKYRTYNVAI EGTSLVTVYR NQFGEIIKAK GIDGLIAELK AKNGKK*

```

ORF91ng-1 and ORF91-1 show 92.3% identity in 196 aa overlap:

	10	20	30	40	50	60
25 orf91-1.pep	MKKSSLISALGIGILSIGMAFAPADAVS0IRQNATQVLSILKNGDANTARQKAEEAYAIP	: : : : : : : : : : :				
orf91ng-1	MKKSSFISALGIGILSIGMAFSPADAVS0IRQNATQVLTILKSGDAASARPKAEEAYAVP	10 20 30 40 50 60				
30 orf91-1.pep	YDFDFORMTALAVGNPWRATASDAQKQALAKEFQTLLIRIYTSGTMLKLNKANVNVKDNPIVN	70 80 90 100 110 120				
orf91ng-1	YDFDFQRMTALAVGNPWRATASDAQKQALAKEFQTLLIRIYTSGTMLKFKNATVNVKDNPIVN	70 80 90 100 110 120				
35 orf91-1.pep	KGGKEIIVRAEVGVPGKPVNMDFTTYQSGGKYRTYNAIEGASLTVTVRNQFGEIILKAK	130 140 150 160 170 180				
orf91ng-1	KGGKEIIVRAEVGVPGKPVNMDFTTYQSGGKYRTYNAIEGTSLVTVRNQFGEIILKAK	130 140 150 160 170 180				
40 orf91-1.pep	GVDGLIAELKAKNGKK	190				
orf91ng-1	GIDGLIAELKAKNGKK	190				

In addition, ORF91ng-1 shows homology to a hypothetical *E.coli* protein:

```

sp|P45390|YRCB_ECOLI HYPOTHETICAL 24.0 KD PROTEIN IN MURZ-RPON INTERGENIC
REGION PRECURSOR (F211) >gi|606130 (U18997) ORF f211 [Escherichia coli]
>gi|1789583 (AE000399) hypothetical 24.0 kd protein in murz-rpon intergenic
region [Escherichia coli] length = 211

Score = 70.6 bits (170), Expect = 6e-12
Identities = 42/137 (30%), Positives = 76/137 (54%), Gaps = 6/137 (4%)
55 Query: 59 VPYDFQRMATLAVGNPWRATASDAQKQALAKEFQTLLIRIYTSGTMLKFNATVNVKDNPI 118
+FY + AL +G +++A+ AQ++A F+ L +Y + + T + P
Sbjct: 65 LPYVQVKYAGALVLGQYYKSATPAQREAYFAAFREYLKQAYGQALAMYHGOTYQIA--PE 122

Query: 119 VPKGGKEIV--VRAEVGVIP-GQKEVNMDFTTYQSG--GKYRTYNVAIEGTSLVTVRNQFQG 174
G K IV + R + P G + FV +DF ++ G ++ Y++ EG S++T +N+G
Sbjct: 123 QPLGDKTIVPIRVTLIDPNGRPPVRLDFQWKRNSQTNWQAYDMIAEGVSMITTKQNEWG 182

```

Query: 175 EIIKAKGIDGLIAELKA 191
 +++ KGIDGL A+LK+
Sbjct: 183 TLLRTKGIDGLTAQLKS 199

Based on this analysis, including the presence of a putative leader sequence in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 42

The following DNA sequence was identified in *N.meningitidis* <SEQ ID 355>:

10	1	ATGAAACACA	TACTCCGG	GATTGCCGCA	TCCGCACTAAC	GTATTCCAGA
	51	CGCTTGCAG	CATCCGCCG	GCGAACCGAA	CATCAAAAC	GAACGCCGTC
	101	TGTGCAACGA	TACCTCCAT	TCAAACATAAC	FTTTTGNNNN	nnnnnnnnnn
	151	nnnnnnnnnn	nnGCACTAA	AAGAACGGGG	ATGGCACATT	TTGGCCGTAT
	201	CGACCATCTAG	GAACGCCG	GCCGAAACGG	CTTAACTGATG	ACGGCCGGA
	251	AAGTCATGCT	CTTGGCAGAC	CCCCAAACGG	ACGGCCGCT	GATGTGCAA
	301	GACCCGGCGCT	TGGCCGCTG	ACTGCGCCCTA	GGCGTCGGC	TATGGCAAC
	351	GGACGCCGAA	GTAGGGCGCA	CCTTACATAG	TAGCGCCGCC	CTCATGGCG
	401	CGACGCCGCA	CGGTTCGCCA	GAAGTGGCCTA	ACATTTGGC	AAACGCCGAA
	451	AAATGCTTAC	AAAACCCACT	AGGGCAATTA		

This corresponds to the amino acid sequence <SEQ ID 356; ORF97>:

20 1 MKHILPLIIA SALCISTASA HPASEPSTQN ETAMITHLTI SKYSGFGXXXX
51 XXXXAIKSKG MDIPAVIDHQ EAARRNLTM QPKAVLVFGT PKAGTPLMVK
101 101 DPAFLAQPL RVLVTETDGK VRAAYTDTRA LIAGSRLGFD EVANTIANAE
151 151 KLIQKTVTG*

Further work revealed the complete nucleotide sequence <SEQ ID 357>;

25		1 ATGAAACACA TACTCCCGCT GATTGCGC GA TCGCACTCT GC ATTTCAC 51 CGCTTGGCGA CATCTGCCTCGA GCGGAAGCGT CACCAAANAC AACGACCGTA 101 TGACCCACC TACCTCCAC TCAAAATATA GTTTTGACGA AAGCGTCAGC 151 CGCGCTTGTAA CGCGCCATAA AAGAACAGGGG ATGGACATTG TTGGCTGTAT 201 CGGACCATCGA GAAGGCCCGC GCGGAACCGG CTAAAGCTG AGCGGCCGCG
	251	AACTCTATCGT CTGGCGGACG CCCAAAGCGG GCACGGCGCT GATGGCTCAA 301 GACCCGGCGT TGCCCCCTGGA ACTGCCCCTA CGGTGCTCGT TTACGAAAGA 351 GGACGGCCAA GTAGCGGCCG CCTATAGCGA TAGCGCCGGC CTCTACCGCG 401 GCAGCGCCAT CGGTGTTCCGA GAGTGGCGCA ACATTTTGGC AAACGCGCGA 451 AAACTGATAC AAAAACCGCT AGGGCAATAA

35 This corresponds to the amino acid sequence <SEQ ID 358; ORF97-1>;

1 MKHILPLIAA SALCISTASA HPASEPSTQN ETAMTHTLT SKYSFDETVS
51 RLETAISKG MDIFAVIDHQ EAARRGLTM QPAKIVFGT PKAGTPLMVK
101 DPAFAQLQLP RVLVTETDGK VRAAYTDTRA LIAGSRIGFD EVANTLANAE
151 KLIQKTVGE*

40 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF97 shows 88.7% identity over a 159aa overlap with an ORF (ORF97a) from strain A of *N. meningitidis*:

```

45      10   20   30   40   50   60
orf97.pep MKHILPLIAASALCISTASAHPASEPSTQNETAMITHLISKYSGFXXXXXXAIKSKG
          ||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :
orf97a  MKHILPLXXASALCISTASXHPASEPQTQNETAMITHLISKYSGFDETVSRLETAIKSKG
          10   20   30   40   50   60

```

-232-

		70	80	90	100	110	120
	orf97.pep	MDIFAVIDHQEAARRNGLTMQPAKIVVFGTPKAGTPLMVKDPAFALQLPLRVLVLTETDGK					
5	orf97a	MDIFAVIDHQEAARRNGLTMQPAKIVVFGTPKAGTPLMVKDPAFALQLPLRVLVLTETDGK	70	80	90	100	110
							120
		130	140	150	160		
	orf97.pep	VRAAYTDTRALIAGSRIGFDEVANTLANAEKLIQKTVGEX					
10	orf97a	VRAAYTDTRALIAGSRIGFDEVANTLANAEKLIQKTVGEX	130	140	150	160	

The complete length ORF97a nucleotide sequence <SEQ ID 359> is:

1	ATGAAACACA	TACTCCCGCT	GATGNGCGA	TCGGACTCT	GCATTTCAC
5	51	CGCTTGGNN	CATCTGCCA	CGAACCGCA	AACCCAAAAC
10	101	TGACCAACGG	TACCCCTACC	TCAAATACA	GTTTGAGCGA
15	151	CGCCCTGAAA	CGGCCATATAA	AAGCAGAAGG	ATGGACATT
20	201	CGGACCATAG	GAAGCCGGCC	GCGGAAAGG	CTTAAACGATG
25	251	AAGTCATGCT	CTTCGGCGCT	CCCAAAGCGG	GTACGCGGCT
30	301	GACCCCGCT	TGGCCCTGCA	ACTGGCCCG	CGCGCTCTTG
35	351	GGAGCCCAA	GTACGGCGCG	CCTATACCGA	TAGGGGGCC
40	401	GCAGCCGCT	CGGTTGCGAC	GAAGTGGCA	ACACITGGC
45	451	AACTGTATAC	AAAAACCAT	AGGGAAATAA	AAACGCCGAA

This encodes a protein having amino acid sequence <SEQ ID 360>:

1	MXHILPLXXA	SALCISTASXY	H ₂ SEAPOTON	EPAAMITHLTLT	SKYSFDETWS
25	51	RLETAIKSKG	MDIFAVIDHQ	BAARRNGLTM	QPAKIVVFGT
	101	DPAFALQLPL	RVXVLTETDGK	VRAAYTDTRA	LIAGSRIGFD
	151	KLIQKTV*			

ORF97a and ORF97-1 show 95.6% identity in 159 aa overlap:

30	orf97a.pep	MXHILPLXXASALCISTASXYH ₂ SEAPOTQNETAMITMTHLTSKYSFDETWSRLETAIKSKG	10	20	30	40	50	60
	orf97-1	MKHILPLIAASALCISTASAH ₂ SEAPSTQNETAMITMTHLTSKYSFDETWSRLETAIKSKG	10	20	30	40	50	60
35	orf97a.pep	MDIFAVIDHQEAARRNGLTMQPAKIVVFGTPKAGTPLMVKDPAFALQLPLRVLVLTETDGK	70	80	90	100	110	120
	orf97-1	MDIFAVIDHQEAARRNGLTMQPAKIVVFGTPKAGTPLMVKDPAFALQLPLRVLVLTETDGK	70	80	90	100	110	120
40	orf97a.pep	VRAAYTDTRALIAGSRIGFDEVANTLANAEKLIQKTVGEX	130	140	150	160		
	orf97-1	VRAAYTDTRALIAGSRIGFDEVANTLANAEKLIQKTVGEX	130	140	150	160		

Homology with a predicted ORF from *N.gonorrhoeae*

ORF97 shows 88.1% identity over a 159aa overlap with a predicted ORF (ORF97.ng) from *N.gonorrhoeae*:

50	orf97.pep	MKHILPLIAASALCISTASAH ₂ SEAPSTQNETAMITMTHLTSKYSFGXXXXXXXXXAIKSKG	60
	orf97ng	MKHILPLIAASACFCISTASAH ₂ PAGKPTQNETAMITMTHLTSKYSFDETWSRLETAIKSKG	60
55	orf97.pep	MDIFAVIDHQEAARRNGLTMQPAKIVVFGTPKAGTPLMVKDPAFALQLPLRVLVLTETDGK	120
	orf97ng	MDIFAVIDHQEAARRNGLTMQPAKIVVFGTPKAGTPLMVKDPAFALQLPLRVLVLTETDGK	120
60	orf97.pep	VRAAYTDTRALIAGSRIGFDEVANTLANAEKLIQKTVGEE	159
	orf97ng	VRTAYTDTRALIVGSRISFDEVANTLANAEKLIQKTVGEE	159

The complete length ORF97ng nucleotide sequence <SEQ ID 361> is predicted to encode a protein having amino acid sequence <SEQ ID 362>;

1 MKHILPPIAA SAFCISTASA HPAGKPTQN ETAMITHTLT SKYSFDETVS
51 RLETAIKSKG MDIFAVIDHQ EAARRNGLTM QPAKIVVFGT PKAGTPLMVK
101 DPFAFLQPL RLVLTETDGK VRTAYTDTRA LIVGSRISFD EVANTLANAE
151 KLTOKTVC*

Further work revealed the complete nucleotide sequence <SEQ ID 363>:

10	1 ATGAAACACA TACTCCCTat gatcgccga TccgcactCT GCATTCAC 51 CGCTTGGCGA CACCTUCGCG GCAAAACGCCGG CACCCAAAAC GAAAGCTGA 101 TGACCCACCA CACCTTCACCC TTGCAATTATA GTTGTGACGC AACCTTGACG 151 CGCCGTTCAA CGGCGCAAAATA AAGCGAAAGGG ATGGACATTG CTGTGACCATTT 201 CGGACCATGAA GAAGCGGCCAG CGCGGAACCGG CTGGACCATG CGCCGGCCA 251 AAGTCATCTG CTTCGGCAGG CGCCAGGCCCG STACCGCCTt GTATGCTAAAC 301 GACCCGGCTT FGCGCTCTGUA ACTGCGCTGTG CGCGCTCTG TTACCGAAGAA 351 GGACGCCCAA TAATGCGACCU CCTATAGCTTA GAGTCGGCGCC ACATCTCG 401 CGACGGCCGAT CAGTTTCGCA GAACTGGCCAA ACATCTTGGC AAACCGCGAA 451 AAATGATGAC AAAACCCCTT AGCGCAATAA
15	

This corresponds to the amino acid sequence <SEQ ID 364; ORF97ng-1>;

20 1 MKHILPLIIA SALCISTASA HPGAKPPTQN ETAMTTHTLT SKYSFDETVS
51 RLETAIKSKG MDTFAVIDHQ EAARRNGLTM QPAKIVFGT PKAGTPLMVK
101 DPAFALQLPL RVLVTETDQK VRTAYTDTRA LIVGSRISFD EVANTLANAE
151 KLIOTVGE*

ORF97ng-1 and ORF97-1 show 96.2% identity in 159 aa overlap:

		10	20	30	40	50	60
25	orf97-1.pep	MKHILPLIRASALCISTASAHPASEPSTQNETAMTHTLTSKYSFDETCSRLETAISKG : : : : : : : : : : : : : orf97ng-1	MKHILPLIRASALCISTASAHPAGKPPTQNETAMTHTLTSKYSFDETCSRLETAISKG 10 20 30 40 50 60				
	orf97-1.pep	70	80	90	100	110	120
30	orf97-1.pep	MDFIAVIDHQEAARRNGLTMQPAKIVFVGTPKAGTPKLVMVKDPAFALQLPRLRVLTETDGK : : : : : : : : : : : : : orf97ng-1	MDFIAVIDHQEAARRNGLTMQPAKIVFVGTPKAGTPKLVMVKDPAFALQLPRLRVLTETDGK 70 80 90 100 110 120				
	orf97-1.pep	130	140	150	160		
35	orf97-1.pep	VRAAYDTTRALIAGSRSIGDEVANTLANAEKLIQKTVGEX : : : : : : : : : : orf97ng-1	VRTAYDTTRALIAGSRSIDFEVANTLANAEKLIQKTVGEX 130 140 150 160				
	orf97-1.pep						
40	orf97-1.pep						
	orf97ng-1						

Based on this analysis, including the presence of a putative leader sequence in the gonococcal protein, it was predicted that the proteins from *N. meningitidis* and *N. gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF97-1 (15.3kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figures 12A & 12B show, respectively, the results of affinity purification of the GST-fusion and His-fusion proteins. Purified GST-fusion protein was used to immunise mice, whose sera were used for Western Blot (Figure 12C), ELISA (positive result), and FACS analysis (Figure 12D). These experiments confirm that ORF97-1 is a surface-exposed protein and that it is a useful immunogen.

Figure 12E shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF97-1.

Example 43

The following DNA, believed to be complete, sequence was identified in *N.meningitidis* <SEQ ID 365>:

```

5      1 ATGGCTTTA TTACGCCTT ATTCAAAGC AGTAAATGGC TGATTGTGCC
51     51 GCTGATGTC CCCGCCTTC AGAATGTGGC GCGGGAGGGG ATAGATGTGA
101    101 GCGGTGCGCA AGCGAGGATA ACCGACGCGC GGCACGCTTC CATCAGCACG
151    151 CGCTTCAAA CGGAGCTGCC CGAACAGCTC CAACAGCGGT TGGCCCGGGg
201    201 CGTGGCGCTC AACITTAACCT TAAGCTGGCC GCTTTCGGCC CGGATATAATCG
251    251 CTTCCTTATCG GTTAAATATT GGCGCAACTG TTGGCGATGA CGACAATATT
301    301 GACTACAAAC TGAGTTCTCA TCCGGCTGACG AAACCGCTACC CGGTATACCGT
351    351 CGCGGGGTT TCAGCACACT ACCGACACCTT GGATCGCGCA TTGGCGCGGA
401    401 CGCGGGGGT TGCCAACTGG AAAGTCTCTGA ACAAAAGGGCGC GCTGTCGGGT
451    451 CGGGGACAG GGGAAACCAA GGGCGRAAATC CGCGTGCAGC TGTCACTTC
501    501 AAAACTGCCA AAGCCTTTT AAATCAATGC ATTGACTTCT CAAAACTGGC
551    551 ATTTGGATTC GGTTGGAAA CCTCTAAACA TCATCGGGAA CAAATAA

```

This corresponds to the amino acid sequence <SEQ ID 366; ORF106>:

```

1 MAFITRLFKS SKWLIVPLML PAFQNVAAEG IDVSRRAEARI TDGGQLSISS
20 20 RFQTELPDQL QOALRRGVPL NFTLWSQLSA PIIASYRFKL GQLIGDDNI
101 101 DYKLSTHPFL KRYRVTVGAF STDYDTLDAA LRATGAVANW KVLNKGALSG
151 151 AEAGETKAEI RLTLSTSKLP KPQINALTNS QNWHLDSGWK PLNIIIGNK*

```

Further work revealed the following DNA sequence <SEQ ID 367>:

```

1 ATGGCTTTA TTACGCCTT ATTCAAAGC AGTAAATGGC TGATTGTGCC
51 GCTGATGTC CCCGCCTTC AGAATGTGGC GCGGGAGGGG ATAGATGTGA
101 GCGGTGCGCA AGCGAGGATA ACCGACGCGC GGCACGCTTC CATCAGCACG
151 CGCTTCAAA CGGAGCTGCC CGAACAGCTC CAACAGCGGT TGGCCCGGGg
201 CGTGGCGCTC AACITTAACCT TAAGCTGGCC GCTTTCGGCC CGGATATAATCG
251 CTTCCTTATCG GTTAAATATT GGCGCAACTG TTGGCGATGA CGACAATATT
301 GACTACAAAC TGAGTTCTCA TCCGGCTGACG AAACCGCTACC CGGTATACCGT
351 CGCGGGGTT TCAGCACACT ACCGACACCTT GGATCGCGCA TTGGCGCGGA
401 CGCGGGGGT TGCCAACTGG AAAGTCTCTGA ACAAAAGGGCGC GCTGTCGGGT
451 CGGGGACAG GGGAAACCAA GGGCGRAAATC CGCGTGCAGC TGTCACTTC
501 AAAACTGCCA AAGCCTTTT AAATCAATGC ATTGACTTCT CAAAACTGGC
551 ATTTGGATTC GGTTGGAAA CCTCTAAACA TCATCGGGAA CAAATAA

```

35 This corresponds to the amino acid sequence <SEQ ID 368; ORF106-1>:

```

1 MAFITRLFKS SKWLIVPLML PAFQNVAAEG IDVSRRAEARI TDGGQLSISS
51 RFQTELPDQL QOALRRGVPL NFTLWSQLSA PIIASYRFKL GQLIGDDNI
101 DYKLSTHPFL KRYRVTVGAF STDYDTLDAA LRATGAVANW KVLNKGALSG
151 AEAGETKAEI RLTLSTSKLP KPQINALTNS QNWHLDSGWK PLNIIIGNK*

```

40 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF106 shows 87.4% identity over a 199aa overlap with an ORF (ORF106a) from strain A of *N.meningitidis*:

45	<code>orfl06.pep</code> MAFITRLFKSSK-WLIVPLMLPAFQNVAAEG IDVSRRAEARI TDGGQLSISSRFQTELPO :: : :: : : : : : : : : : : : : : <code>orfl06a</code> MAFITRLFKSIRQWLVLPMISVLFDAAEAGIDVSRRAEARI DGGQLSXSRFQTELPO 10 20 30 40 50 60	10 20 30 40 50 59
50	<code>orfl06.pep</code> LQQALRRGVPLNFTLWSQLSAPIIASYRFKLGQLIGDDNIYKLSFHPLTKRYRVTVG : : : : : : : : : : : : : : : : 60 70 80 90 100 110 119	60

-235-

	orf106a	LQXAXXRGVXLNXTLXWQLSAPIIASYRFXLSQLIGDDDXIDYKLSFHPLTNRYRVTVGA 70 80 90 100 110 120
5	orf106.pep	FSTDYDTLDAAALRATGAVANWKVLNKGALS 120 130 140 150 160 170 179 FSTXYDTLDAAALRATGAVANWKVLNKGALS 130 140 150 160 170 180
10	orf106a	SQNWHLDSGWKPLNIIIGNKX 180 190 199 SQNWHLDSGWKPLNIIIGNKX 190 200
15	orf106.pep	ATGGCTTTTA TTAGCCGCTT ATTCAAAGC ATTARACAAAT GGCTTGTC 51 GCTCGCGATG CTTCGCGTTT TGCGGAGC GGGGGCGGAG GGGATAGATG 101 TGAGCCGGGC CGAACGAGG ATAAANGAGC GCGGGCACGT TTCCATNAGN 151 AGCCGCTTCA AAACAGGACT CGCCGACCCAGCT CCTCAAANNNG CGNNNGNGCG 201 GGGCGTGNCC CTCACAINTA CCTTTAAGNTIC GCAGCTTTCC CCCCCGATAA 251 TCGGCTCTTA TCCTGGTTTAA TTGGGCCAAC TGATTTGGCGA TGAGCAGCAT 301 ATTGAGTACAA AACITGAGTTT CCATCCGGCTG ACCAAACGGCT ACCCGCTTAC 351 CGTCGGCGCGG TTTCGGACAGC ANTACGACAC CTGGGATGGCG SCATTGCGCG 401 CGACCGCGCGG GGTTCGCAAC TGAAAAGTC TGACAAAAGG CGGGCTGTGCG 451 GGTGCGCGAAC CGGGGAAAC CAAGGGGAAAT ATCCGCGCTGA CGCTGTGCCAC 501 TCTAAACTCTG CCCAACGCTT TTCAAATCAA TGCAATTGACT TCTCAAAACT 551 GGCATTGGAA TTCCGGTTGG AAACCTCTAA ACATCATCGG GAACAAATAA

- 30 Due to the K→N substitution at residue 111, the homology between ORF106a and ORF106-1 is 87.9% over the same 199 aa overlap.

The complete length ORF106a nucleotide sequence <SEQ ID 369> is:

1	MAFITRLFKS IKQWLIVLPPM LSVLPDAAAEG GIDVSRAEARITDGGQLSISSRFQTEL 51 SRFQTELPGQ LQXAXXRGVX LNXTLWQLS APIIASYRFXL LGQLIGDDDX 101 IDYKLSPHML TNRYRVTVGA FSTXYDTLDAA LRATGAVAN WKVLNKGALS 151 GAEAGETKAE IRLTISTSKL PKPFQINALT SQNWHLDSGW KPLNIIIGNK*
20	
25	
30	This encodes a protein having amino acid sequence <SEQ ID 370>:
35	

1	MAFITRLFKS IKQWLIVLPPM LSVLPDAAAEG GIDVSRAEARITDGGQLSISSRFQTEL 51 SRFQTELPGQ LQXAXXRGVX LNXTLWQLS APIIASYRFXL LGQLIGDDDX 101 IDYKLSPHML TNRYRVTVGA FSTXYDTLDAA LRATGAVAN WKVLNKGALS 151 GAEAGETKAE IRLTISTSKL PKPFQINALT SQNWHLDSGW KPLNIIIGNK*
20	
25	
30	
35	

Homology with a predicted ORF from *N.gonorrhoeae*

ORF106 shows 90.5% identity over a 199aa overlap with a predicted ORF (ORF106.ng) from *N.gonorrhoeae*:

40	orf106.pep	MAFITRLPKSSK -WLIVPLNLPAFONVAEAGIDVSRAEARITDGGQLSISSRFQTEL orf106ng	MAFITRLFKS IKQWLIVLPPM LSVLPDAAAEG GIDVSRAEARITDGGQLSISSRFQTEL 59 60
45	orf106.pep	LGQALLRGVRPELNFTLSWQLSAPIIASYRFXLSQLIGDDDNIDYKLSFHPLTKRYRVT orf106ng	LGQALLRGVRPELNFTLSWQLSAPIIASYRFXLSQLIGDDDNIDYKLSFHPLTKRYRVT 119 120
50	orf106.pep	FSTDYDTLDAAALRATGAVANWKVLNKGALS orf106ng	FSTDYDTLDAAALRATGAVANWKVLNKGALS 179 180
55	orf106.pep	SQNWHLDSGWKPLNIIIGNKX orf106ng	SQNWHLDSGWKPLNIIIGNKX 198 SQNWHLDSGWKPLNIIIGNKX 199

Due to the K→N substitution at residue 111, the homology between ORF106ng and ORF106-1 is 91.0% over the same 199 aa overlap.

The complete length ORF106ng nucleotide sequence <SEQ ID 371> is:

```

1 ATGGGCTTTA TTACCGCCTT ATTCAAAGC ATTAACAAAT GGCTTGTCGT
51 GTTCCCGATA CTCTCCGTT TGCGGAGCG GGCAGCGGAG GGCAATTGCGG
101 CAACCGCGCG CGAACGAGG ATAACCGACG GCGGGCGGAG TTCCATCAGC
151 ACCCGCTTCC AAACCGAGCT GCGCGACCCAG CTCCAAACAGG CGTTGCGCCG
201 GGGCGTACCC CTCAACTTAA CCTTAAAGCTG GCAGCTTTCG GCCCCGACAA
251 TCGCTTCTTA TCGGTTTAAAT TTGGGGCACAC TGATTTGGCGA TGACGACAT
301 ATGACTACAA AACTAAGTTT CCATCGCGTGC ACCAACCGCT ACCCGCTTAC
351 CGTCGGCGCG TTTCACACCG ATTACGACAC TTTGGATGGG GCATTGCGG
401 CGACCGCGCG GTTGGCCACG TGAAAAGTCC TGAAACAAAGG CGCGTGTGCC
451 GTGCGGACAG CAGGGGAAAC AGAACGGGAA ATCCGCGTGA CGCTGTCAC
501 TTCAAAACTCA CCCAACGCTT TCCAATCTAA CCCATGACT TCTCAAAACT
551 GGCAATTGCA TTCCGGTGTG AAACCTCTAA ACATCATCGG GAACANATAA

```

This encodes a protein having amino acid sequence <SEQ ID 372>:

```

15 1 MAFITRLFKS IKQWLVLPI LSVLPDAAAE GIAATRAEAR IITDGGRLSIS
51 5 SRFQTELPDQ LQLQRREGVP LNFTLWSWOLN APTIASYRKF LGQLIGDDDN
101 10 IDYKLFHPL THRYRVTVGA FSTDYDTLDA ALRATGAVAN WKVILNKGALS
151 15 GAEGEGETRAE IRLTLSSTKL PKFFQINALT SQNHLDGSW KPLNIIGNK*

```

Based on this analysis, including the presence of a putative leader sequence in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF106-1 (18kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 13A shows the results of affinity purification of the His-fusion protein, and Figure 13B shows the results of expression of the GST-fusion in *E.coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for FACS analysis (Figure 13C). These experiments confirm that ORF106-1 is a surface-exposed protein, and that it is a useful immunogen.

Example 44

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID

30 373>:

```

1 ATGGACACAA AAGAAATCCT CGG TACCGG GcAGGGcTCGA TGCGGAGCGC
51 GGTTTTCAGCC GTCATCATCC TGCGCGTGTG GTCTGTTATG TTCCCCCGCG
101 ACGACATCGG CGCGCTCGTG CTGATCGAGA CGCGCGCGGG GCGTACGGTG
151 TCGGTGTTGT GCTCGGGGGT GGATCGGCA TAATCGCGGG AATACTATGCG
201 CACCGCCGAA AAAGACACCT TGTTCAAAAC CCTGTTCTCG CGCGCGCTGC
251 TGCTCGCGC CGCGATAGCC GCGCTCTGCG TTTCGGCGCC CGCCCTCGCG
301 TCTGAATTCG TGTTCGACT CGAGATGCG qCGCGCGUGGq TCAGGGCTGGT
351 GCTGTTTGAA CTGAGCTTCC CGGCCATCCC cTTTCTTTTA CTGTTTTTGC
401 GTATGGAAAGC AGCGCCGCTT GCGTTTTGCG CGCGGCAACT CGTGCcCAAG
451 CTCGGCCATCC TGCCTGCTG_T GCGCGCGAGC GTCCGGCGTGC FGCACTTTCC
501 AGCGAAACACA CGCGCTCGTG CGCGCGTTTA CGCGCTGGCA AACCTTGGCG
551 CGCGCGCTTG TTGCTGTTT CAAACCGAT CGCGCTGGAA CGCGCGCGCG
601 CACCGACCGG TTTCGGCGCG CGTCATGCGAC CGGGGG..TGC GCTACCGGAT
651 ACCGGATCGCA CTGAGCGAGCA TGCGCTATTG GGCGCTGGCA TCACCGCGACC
701 GTTGTGTTCT GAAAATATA CGCGCGCTTGG AACACGCTGG CGTTTATTCG
751 ATGGGCTATTG CGTGTGGGG GCGCGCATTA TTGTTTCAAA CGATCTTTTC
801 AACGGCTGG ACACCGTATA TTTCGGCGCA ATTCGAAGAA AACGCCCGC
851 CGCGCTGGCT CGCGCGACAG CGAGATCGG CGCGCGCCCT GCTTGGCTTC
901 GCGCTCTGC_ TGACGGCGAT TTTCGCGCC TTGCGCTCCC TCTGCTGCC
951 GGAAACATAC CGCGCGCTCC GGTTTATCGT CGTATCGT ATG..TGCGCG

```

1001	CGCTGTGTTG	CACGCTGGGG	GAAATCAGCG	GCATCGGTTT	GAACGTCGTT
1051	CGCAAAACGC	CCGGCGATCG	GCTGCCAAC	TTGGCGSCGC	TGGCGGCAA
1101	CTCTGCTGCTG	CTGGGGCTTG	ACCGTSCCGT	ACCGGCGGAGG	CGCGC, GCGG
1151	CGCGCGTTGC	CTGTGCGCCG	TCATTCTGCG	TSTTTTTTGG	CTTCAGAGCC
5	1201	GAAAAGCTyT	GGCGCGCTG	GCAGCGCTC	AAACGCGCTG
	1251	GCACACATGG	TTCTGCTGTA	CTCTCTCGC	GGCCATACACC
	1301	CGCGGCAA	CTATCCCCTG	TTTGCGGCGC	TATGGCGGCA
	1351	GGCTGCATCC	TSGCCGACCG	GAAAGATTTG	CACAAACTGT
	1401	AAAAAAACAA	GGTTTCCCAT	TATGA	

10 This corresponds to the amino acid sequence <SEQ ID 374; ORF10>:

1	MDTKEILYXA	AGSISGVLA	VIIPLLSWY	FPAADDIGRIV	LMQTAAGLTV
5	51	SVLCLGLDQA	YUREYATAD	KDTLFKTLFL	PPLLSAAAIA
	101	SEILFLSLDDA	AAGIGLVLF	LSFLPLIRFL	LVLRMEGRAL
	151	LAILLXXPL	VGLLHFPANT	AVLTAVYALA	NLAAAFLLF
	201	HAFPSFPAVL	RGRYRIPVIA	LSSIAIWGLA	SADRFLFKY
	251	MGISPSFGAL	LFQSFIFSTW	TPYIPIRAIEE	AGLEQLGVYS
	301	ALCTGTIFPS	LASLILEPNY	AAVRFIVVVS	NAPPARLISAT
	351	RKTRPIALAT	LGALAANLLL	MGLDRAVPAP	AESAAALLAS
20	401	ESSCRWLQPL	KRLPYLHHTL	FCLTSSAAYT	EISIGLGNVV
	451	GCILRHRKDL	HKLHFYLKKQ	CGFTPANYPL	SWLWFFAKT

Further sequence analysis revealed the complete DNA sequence<SEQ ID 375> to be:

1	ATGGACACAA	AAGAAATCCT	CGGCTACCGG	GCAGGCTCGA	TCGGCAGCGC
5	51	GGTTFATGCC	GTCTCATCC	TGGCGCTGCT	GTCTGGGTAT
	101	ACCGACATGG	GGCCATCGG	CTGATCGAGA	GGCGCGGGGG
	151	TCCGTGTTG	GGCTCGGGCT	GGATCAGCGA	TACGTCGGGG
	201	CACCGCCGAC	AAAGACACCT	TGTTCAAACAC	AATACTATATGC
	251	TGTCGTCCG	CACGATAGCC	GGCTCTCGC	CTGTGTCCTG
	301	TCTGAAATTC	TGTTTTACT	CGACGATGSC	GGCGCGCGCA
	351	TCGTTGTTGAC	CTGGACTCTC	TGCCCCATCG	CTTCTCTTA
30	401	GTATGGAAGC	ACGGCCCTT	GGCTTTTGTG	CTGGGCTTAC
	451	CTCGGCCATCC	TGCTGCTGTA	GGCGCTGACT	TCGCGCTGTC
	501	AGCGAACACCC	GGCGCTCTGA	CCGGCGTTTA	CCCGCTGCGCA
	551	CCCGCCCTT	TTCGCTGTTT	CAAAACGGAA	GGCGCTGAA
	601	CACCGACCGG	TTTGCCTGGCC	GGCTCTGCGA	GGCGCTGCGG
	651	ACCGATGCCA	CTGAGCACGA	TCGGCTATTG	GGGGCTGCGA
	701	GTGTTTGTCT	GGAAATATAT	GGCGCGCTTG	TCCCGCGAC
	751	ATGGGTAATT	CTGGCGCCG	AAACGCTCGG	CGTITATTG
	801	AAACGGCTGG	ACACCGPAT	TTTTCGGCG	ATATCGAAAGA
	851	CCCGCCGCGT	CTCGCAACG	CGCAAAATCGG	AAACGCCCGCG
40	901	GCCCTCTG	TGACCGCGCAT	TTCTGCGCC	CTTGCCTCCC
	951	GGAAAACACT	GGCCGCGCTCC	GGTTTATCTG	TCTGCTGTGC
	1001	CGCTGTGTTG	CGCTGCGGG	CGATCAGCGG	GAACGTCGTC
	1051	CGGAAACACG	GGCCGATCGC	GCTGCCAAC	TTGGCGCGC
	1101	CTCTGCTGCTG	CTGGGGCTTG	CGTGGCGTC	GGCGCGCGG
	1151	CGGTGCTGCT	TCTGCGCTTG	TTTTCGGCTT	CAAGACOGAA
45	1201	AGCTTCTGCC	GGCTGTGCGA	GGCGCTCAC	CGCGCTGCCG
	1251	CACATGTTG	TGCTGACCT	CTCGCGCGC	CTACACCTGC
	1301	CGCGAACACTA	TCCCTCTTT	GGCGCGGTAT	TTCGCGACGC
	1351	TGCTACTCTG	GGCCCGGGAA	AGATTTGCGAC	AAACTGTTTC
50	1401	AAAAACAGGT	TTCCTCATAT	GA	ATTATTGAA

This corresponds to the amino acid sequence <SEQ ID 376; ORF10-1>:

1	MDTKEILYXA	AGSISGVLA	VIIPLLSWY	FPAADDIGRIV	LMQTAAGLTV
5	51	SVLCLGLDQA	YUREYATAD	KDTLFKTLFL	PPLLSAAAIA
	101	SEILFLSLDDA	AAGIGLVLF	LSFLPLIRFL	LVLRMEGRAL
	151	LAILLXXPL	VGLLHFPANT	AVLTAVYALA	NLAAAFLLF
	201	HAFPSFPAVL	RGRYRIPVIA	LSSIAIWGLA	SADRFLFKY
	251	MGISPSFGAL	LFQSFIFSTW	TPYIPIRAIEE	AGLEQLGVYS
	301	ALCTGTIFPS	LASLILEPNY	AAVRFIVVVS	NAPPARLISAT
	351	RKTRPIALAT	LGALAANLLL	MGLDRAVPAP	AESAAALLAS
60	401	ESSCRWLQPL	KRLPYLHHTL	FCLTSSAAYT	EISIGLGNVV
	451	CILRHRKDL	HKLHFYLKKQ	CGFTPANYPL	SWLWFFAKT

Computer analysis of this amino acid sequence gave the following results:

Prediction

ORF10-1 is predicted to be the precursor of an integral membrane protein, since it comprises several (12-13) potential transmembrane segments, and a probable cleavable signal peptide

Homology with EpsM from *Streptococcus thermophilus* (accession number U40830).

- 5 ORF10 shows homology with the *epsM* gene of *S. thermophilus*, which encodes a protein of a size similar to ORF10 and is involved in expolysaccharide synthesis. Other homologies are with prokaryotic membrane proteins:

Identities = (25%)

10 Query: 213 LRYGIPLALSSLAYWGLASADRFLKKYAGLEQLGVYSMGIFSFGGAALLLQSIFSTVW 270
 L Y +PL SS + L ++ R F+ + G G++ + +ITF+ W
Sbjct: 210 LYALPPLISISSLWLNNASSRYYFLFLGGAGANGLLAVATKIPSIISIFTFWQW 267

15 Identities = 15/57 (26%), Positives = 31/57 (54%)

Query: 7 LGYAAAGSIGSAVLAVIILPLLSWYFPADDIGRIVLQMTAAGLTVSVLCLGLDQAYVR 63
L + G+GS L + ++PL + + G L QT A L + ++ + A + R
Sbjct: 12 LVFTGTLNGLSKLLVFLVLPYATMPQEYGMADLYQTANNLPLLTIMVNFDATLR 68

20 Identities = 16/96 (16%), Positives = 36/96 (37%)

Homology with a predicted ORF from *N. meningitidis* (strain A)

ORF10 shows 95.4% identity over a 475aa overlap with an ORF (ORF10a) from strain A of *N. meningitidis*:

	10	20	30	40	50	60
orf10.pep	MDTKEILXYAAGSIGSAVLAVIILPLLSWYFPADDIGRIVLMQTAAGLTVSVLCLCLDQA					
orf10a	MDTKEILGYAAGSIGSAVLAVIILPLLSWYFPADDIGRIVLMQTAAGLTVSVLCLGLDQA					
	10	20	30	40	50	60

	70	80	90	100	110	120
	130	140	150	160	170	180
orf10.pep	L	S	F	P	I	R

45	orf10a	LSFLPIRFLLLVLRMGERALAFSSAQLVSKLAILLPLLPTVGLLHFPTANTAVLTAVALA 130 140 150 160 170 180
50	orf10.pep	190 200 210 220 230 240 NLAAGAFLLQFNQRCLRAKVRHAPFSPAVLHRGXRGYIPLASSIAYWGLASADRLEFLKKY

50 orf10a NLAAAFLLEQNRCRLKAVRRAFPSSAVLHRLGRYGIPIALSSIAYWGLASADRLFLKKY
190 200 210 220 230 240
250 260 270 280 290 300

55	orf10.pep	AGLEQLGVYSSMGCSFGGAALLFQSISFTWTPYIFRAIEENAPPARLSATAESAAALLAS
	orf10a	AGLEQLGVYSSMGCSFGGAALLFQSISFTWTPYIFRAIEENAPPARLSATAESAAALLAS

-239-

		310	320	330	340	350	360
	orf10.pep	ALCTXGIFSPLASLILLPENYAAVRFIVVSCMXPPLFCTLAZISIGILNVLVRKTRPIALI					
5	orf10a	ALCLTGIFSPLASLILLPENYAAVRFIVVSCMXPPLFCTLWEISIGILNVLVRKTRPIALI					
		310	320	330	340	350	360
		370	380	390	400	410	419
	orf10.pep	LGALANLNLGLDRAVPAR-FXGAAGAACASFWLFFAFTTESSCRWLQPLKRRLPLYHHT					
10	orf10a	LGALANLNLGLL--AVPSGGARGAACAAACASFWLFFVFKTESSCRWLQPLKRRLPLYHMT					
		370	380	390	400	410	
		420	430	440	450	460	470
	orf10.pep	LFCLTSSAAAYTCFGTPANYPLFAGWVWAYLAGCILRHKRDHLKLHFYLKKQGFPLX					
15	orf10a	LFCLASSAAAYTCFGTPANYPLFAGWVWAYLAGCILRHKRDHLKLHFYLKKQGFPLX					
		420	430	440	450	460	470

The complete length ORF10a nucleotide sequence <SEQ ID 377> is:

20	1	ATGGACACAA	AAGAAATCCT	CGGCTACGCG	CGAGGCTCGA	TCCGAGCGC
	51	GTTGTATTCG	GGCATCATCC	TGCCTGCTGT	TGCTGTTGAT	TTTCTTCGGC
	101	ACGACATCGC	GGCGATCTGTC	CTGTAGCAGA	CGGCCGGCGC	GTCAGCTGTC
	151	TCTGGTTGT	GGCTCTGGG	GGATCAAGCGA	TAAGCTCCCG	ATAACTATGC
	201	CGCCCGCCAC	AAAGAACACTT	TGTTAACAAAC	CTGGCTTCCT	CGCCGGCTGC
	251	TGTCGCGGC	GGGGATAGATG	GGCTCTGTCG	TTTCCCAGGCC	ATTATCTGGC
25	301	TCTGAATCTC	TGGTTCCTG	CGACAGATGCC	CGGCCGGCGC	TGCGGGCTGT
	351	GCTGTGTTGA	TCTGAGCTTC	TGCCGATTCGG	CTTCTCTTCA	CTGTTGTTTC
	401	GTATGGGAGG	ACCGGGCCCTT	GCTTCTTCTG	CGGCCGAACAT	GTGGCTTCAA
	451	CTCGGCATTC	TGCTGCTGCT	GGCGCTGAGG	GTGGGGCTGC	TGCACTTTCC
30	501	GGGCAACACC	GGCGCTCTGG	CGCGCTGTTA	CGCGCTGCGA	AACCTTGGCG
	551	CGGCCGCTT	TTTGCTGTTT	AAAAACCGAT	GGCGCTGCGA	GGCGCTCCGG
	601	CGGGCACCGT	TTTCATCGCC	CGCTCTGCTAT	CGGGCGCTGC	TGACAGCGCC
	651	ACGGATTCG	TAAGCAGACA	TGGCTTATGG	GGGGCTGGCA	TCCCGGACAC
	701	GTTTGTTCT	AAAARATAA	GGCCGCTTGA	AACAGCTCGG	TTCTATTGCG
35	751	ATGGTTATT	CTGGCGGGC	AGCGGCTTAA	TATGGTCCTAA	GAATCTTTTC
	801	AACGGCTCTGG	ACACGGTAT	TTTCTGGCGC	ATTCAGAACG	AACGGCCCGC
	851	CGGCCGCTT	CTGGCGAACG	CGAGAGATCC	CGGCCGCTT	GTCCTGCTCC
	901	GCCCTCTGCC	TAAGCGGCC	TTTCTGGCGC	CTGGCTCTCC	TCTGTGCTC
	951	GGAAACATAC	GGCGCGCTGC	GGTTTTATGT	CTGATGCTG	ATGCTGCTC
40	1001	CCTGCTTTTG	CGACGGTGG	GAATACCGGG	GCATCGGT	GAAGCTGCTC
	1051	CGAAAACAC	GGGGATCTGG	GTCTGGCAC	TTGGGGCGC	TGGGGCGAAC
	1101	CTCTGCTG	CTGGCGCTTG	CGTACCGCT	CGGCCGGCGC	CGCCGGCGCG
	1151	CGGTGCTG	TGGCGCTCC	TTTTGGCTGT	TTTTGGTTT	CAAGACGCG
	1201	AGCTCTGC	GGCTTGTTGC	GGCGCTCAA	CGCGCTGCCG	TTTATATGCA
45	1251	CACATGTC	TGGCTGGG	CTCGGGGGC	TAACCTGTC	TGGCGCCTC
	1301	CGGCCAAT	CCCCCTGTTT	GGCGCGTAT	GGCGCTATA	TCTGGCAGGC
	1351	TGCTCATCG	GGCACCGGG	GAATTGCGAC	AAACTGTTTC	ATTATTGAA
	1401	AAAACACGT	TTCCTATATT	GA		

This encodes a protein having amino acid sequence <SEQ ID 378>.

50	1 MDTKILIGYA AGSISGVA 51 SVALCILGDA YURBYYAADD 101 SEILFSLDDA AAGIGLIVL 151 LAIILPLPTE VGLLHHPPAN 201 RAPFSSAVLH RGLRYGPIIA 251 MGISGGAAL LFQSFISFWN 301 ALCLTGFST LAAAGLAWA 351 RSTRPIADT LGAALMADL 401 SSCLRWLQFLC RPLYMPHTL	VIIPLLSWY FPDADDIGV ALLSRSPLF PLLSLLPLF LVLRMGEFL AFSSAQVLSK NLAAAAFLFL QNRLRKLAIV SADRFLKKY AGLEODLVGS NAFPARSLAT LFSPARSLAT MLPEFLCTSY EISGIGLNVV AFRIVPEEV MLPEFLCTSY LALMADL GLAVUSSGGA RGAVACRAS CLASSARYST EGTPANYPLE AGWVAVYL
55		

QRE10a and QRE10-1 show 95.4% identity in 475 aa overlap:

	10	20	30	40	50	60
orfl0-1.pep	MDTKEILXYAAGSIGSAVLAVIILPLLSWYPADDIGRIVLQMOTAAGLTVSVLCLGLDQA					
orfl0a	MDTKEILXYAAGSIGSAVLAVIILPLLSWYPADDIGRIVLQMOTAAGLTVSVLCLGLDQA					

-240-

		370	380	390	400	410
5	orf10ng pep	LGALAA	NLLLGL	--AVPSGGT	RGAAC	VACASFWL
	orf10nm	LGALAA	NLLLGL	DRAVEFAR	-PXA	GAWACA
10	orf10ng pep	420	430	440	450	460
	orf10nm	LFCL	ASSA	YCFGT	PANYPL	FAGWAA
15	orf10ng pep	101	151	201	251	301
	orf10nm	151	201	251	301	351
20	orf10ng pep	351	401	451	501	551
	orf10nm	401	451	501	551	601
25	orf10ng pep	601	651	701	751	801
	orf10nm	651	701	751	801	851
30	orf10ng pep	851	901	951	1001	1051
	orf10nm	901	951	1001	1051	1101
35	orf10ng pep	1101	1151	1201	1251	1301
	orf10nm	1151	1201	1251	1301	1351
40	orf10ng pep	1351	1401			
	orf10nm	1401				

The complete length ORF10ng nucleotide sequence <SEQ ID 379> is:

1	ATGAGCACAA	AAGAACATCT	CGGCATACGG	CGAACCTCGA	TCCGGCAGCG
51	GTTTTTAGCC	GTCATCATCC	TGCCGCTGCT	TGCTGTTGTT	TTCccccCGG
101	ACGACATCGG	GGCGCATCGG	CTGATGCAGA	CGCGCGCGGG	ACTGACGGTG
151	TGGATTATGT	TGCTGGGGCT	GGATCAGGCA	TAAGTCGGCG	AATACTATGC
201	CGCGCGGAC	AAAGACACT	TGTCAACAA	CTCTGTCTCG	CGCGCGCTGC
251	TGTGTTCCCG	CGCGATAGCC	GGCTCTCGTC	TTTCCCAGCCC	GTCCTCTCGCG
301	TCTGAAATCC	TGTGTTCCCG	CGACATGTCG	GGCGCGCGCA	TGGGGCTGGT
351	GCTGTTGAA	CTGAGCTTC	TGCCCATCTC	CCTTCTCTTA	CTGTTTTTCG
401	GTATGGAA	GGCGCCCGCT	GCTTCTTCTGT	CGCGCAACT	CTGCCCCAAA
451	CTCGGCAATC	TGCTGCTGTT	CGCGCTGACG	TGTCGGCTCG	TGCACTTTCC
501	GGGGAAACACC	TCCGCTCTGA	CGCGCTTTA	CGCGCTGGCA	ACACCTGGCG
551	CCGGCCGCTT	TTTGTCTTT	AAACACCGAT	GGCGCTGTGAA	GGCGCTGCCG
601	CGGGCGCGT	TTTGTCCCGC	CGTCCTGCAC	CGGGGGCTGC	GTCACGGCAT
651	ACCGCTCGCA	CTGAGCACGC	TGCGCTATTG	GGGGCTGCGCA	TCCGGCGAAC
701	GTTTGTCTCT	AAAAAAATAT	GGCGGCATTA	TTGCTCCAAA	GCATCTTTTC
751	ATGGTGTATT	CTGGTGGCGG	GGCGGCATTA	TTGCTCCAAA	GCATCTTTTC
801	AAACGGCTGG	ACACGGTATA	TTTTCGGTGA	AAATCCAAGAA	AAACCGAACGC
851	CCGGCCGCTG	CTCGGCAACG	CGAAATTCG	CGCGCGCCCT	GCTTGGCTCC
901	GCGCTCTGGC	TGACGGGAT	TTTCGCGCC	CTCGCTCTCC	TCTGCTGCG
951	GGAAAACTAC	CGCGCTCC	GGTTTACCGT	CGTATCGTGT	ATGCTGccgc
1001	cgctgtttta	CAAGCTGACG	GAATTCAGCG	CGATCGGTTT	GAACGTGTC
1051	CGCAAAACCG	GTCCGATCGC	GCTTCGACAC	TTGGGGCGGC	TGGGGCAAA
1101	CTCGTCTGCT	CTGGGGCTTG	CGTACCGCT	CGCGCGCACG	CGCGGGCGCG
1151	CGGTGTCGCT	TGCCGCTCTA	CTCTGTTGTC	TTTTTGTTT	CAAGACAGAA
1201	AGCTCTCTGC	GGCTTGGCA	SCCGCTCAA	CGCTGCGCG	TTTATATGCA
1251	CACATGTTG	TGCGCTGCT	CTCGGCGCG	CTACACCTGC	TTCGGCACAC
1301	CGGAAACACTA	CCCCGTTG	ggccggcgtAT	GGGGCGCATA	TCTGGCAGGC
1351	TGGCATCTCG	GCCACCGGAA	AAATTGGCAC	AAACCTGTTT	ATTATTTGA
1401	AAAAACAGGT	TTCCCATATT	GA		

This encodes a protein having amino acid sequence <SEQ ID 380>:

1	MDTKEILGYA	AGSIGSAVLA	VIILPILSWY	FPADDIGRIV	LMQTAAGLTV
51	SVLVLGLDQA	YVREYYAAAD	KDTLFKTLFL	PPPLLSSAAIA	ALLSLRPSLP
101	SEIISLDDA	AAAGIGLVLFE	LSFLPFLRFLL	LVRMEGRAL	AFSSAQLVPK
151	LATLILLPLF	VGLHLPFANT	SVLTAVYALIA	NLAAAAFLIF	QNRCRKLAVR
201	RAPFSPAVLH	RGLRYGIPFLA	LSSLIAYWGILA	SADRFLKLYK	AGLEQLGVYS
251	MGISFGGAA	TSTVW	AAVIFTFAIEE	NATPARLSAT	AEEAAALIAS
301	ALCITCTGSP	LASLLENPEY	AAVIFTVVS	MLPPLFTYL	EISIGLGNV
351	RKTRPTIALAT	LGALANEPNL	LGALAVPSGGT	RGAAVACAA	FWLFVFVKT
401	SSCRWLQPLK	RPLPYMTLFL	CLASSAA	YCFGT	AGVWAAYLAG
451	CILRHKNLH	KLFHYLKKQG	FPL*		

ORF10ng and ORF10-1 show 96.4% identity in 473 aa overlap:

		10	20	30	40	50	60
55	orf10-1.pep	MDTKEILGYA	AGSIGSAVLA	VIILPILSWY	FPADDIGRIV	LMQTAAGLTV	
	orf10ng-1	MDTKEILGYA	AGSIGSAVLA	VIILPILSWY	FPADDIGRIV	LMQTAAGLTV	
60	orf10-1.pep	70	80	90	100	110	120
	orf10ng-1	YVREYYAAAD	KDTLFKTLFL	PPPLLSSAAIA	ALLSLRPSLP	SEIISLDDA	AGIGLVLFE
65	orf10-1.pep	YVREYYAAAD	KDTLFKTLFL	PPPLLSSAAIA	ALLSLRPSLP	SEIISLDDA	AGIGLVLFE
	orf10ng-1	70	80	90	100	110	120

		130	140	150	160	170	180
5	orf10-1.pep	LSFLPIRF ^{LLVLRMEGRALAFSSAQ} LPKLA ^{IL} LLLPLTVGLLHF ^{PANTAVLTAVYALA}					
	orf10ng-1	LSFLPIRF ^{LLVLRMEGRALAFSSAQ} LPKLA ^{IL} LLLPLTVGLLHF ^{PANTAVLTAVYALA}	130	140	150	160	170
10	orf10-1.pep	NLAAA ^{AFFL} QNCRCLKAVRH ^{AFFSPAVLHRGLRYGIP} I ^{ALASSIAYNGLASADRLFLKKY}	190	200	210	220	230
	orf10ng-1	NLAAA ^{AFFL} QNCRCLKAVR ^{RAFFSPAVLHRGLRYGIP} I ^{ALASSIAYNGLASADRLFLKKY}	190	200	210	220	230
15	orf10-1.pep	AGLEQLGVYSMGISFGGA ^{LLFQSISFTSTWTPYIFRAIEENAPPARLSATAESAALLAS}	250	260	270	280	290
	orf10ng-1	AGLEQLGVYSMGISFGGA ^{LLQSISFTSTWTPYIFRAIEENATPARLSATAESAALLAS}	250	260	270	280	290
20	orf10-1.pep	ALCLTGIFSP ^{LASLLL} PENYA ^{AVRFTIVS} SCMLP ^L FCT ^T LAE ^{ISIGIG} LNVRKTRFIALAT	310	320	330	340	350
	orf10ng-1	ALCLTGIFSP ^{LASLLL} PENYA ^{AVRFTIVS} SCMLP ^L FYT ^T TEISIGIGLNVRKTRFIALAT	310	320	330	340	350
25	orf10-1.pep	LGALAAN ^{LLLLL} GLAVPSGGARGAA ^{VACAS} FWLFFAKFTES ^S CR ^L WQFLKR ^L PLYLMTLF	370	380	390	400	410
	orf10ng-1	LGALAAN ^{LLLLL} GLAVPSGG ^T RGA ^{AV} CA ^{AS} FWLFFVFKT ^E SSCR ^L WQFLKR ^L PLYLMTLF	370	380	390	400	410
30	orf10-1.pep	CLTSSAA ^{YCFGT} PANYPL ^{FAGV} WAAYLAGC ^I L ^R HKD ^L H ^K L ^F H ^Y LKKQGFPLX	430	440	450	460	470
	orf10ng-1	CLASSAA ^{YCFGT} PANYPL ^{FAGV} WAAYLAGC ^I L ^R HKN ^L H ^K L ^F H ^Y LKKQGFPLX	430	440	450	460	470

Based on this analysis, including the presence of a putative leader peptide and several transmembrane segments and the presence of a leucine-zipper motif (4 Leu residues spaced by 6 aa, shown in bold), it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 45

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 381>:

45	1..ATCTCGAAC CGCATAAACCA GCTTAAGGAA GACATCCAACT CTGATCCGC
	51 CGATCAAAC CGCTTGTGCC AACCGGATCGC TCGCGACAG GCAGAGCAGT
	101 CCGATCGGGA AAATGCTGCC GCAAGCAGC CGCTTGC ^G GA TAAAGCCGAC
	151 GAGGTGAAG AAAAGCGGG CGAGCGGAA CGGGAGAGC CGAGCGGACA
	201 GGCAGTGGCT AAGAACGCC TGACGGAGA ^G CGCTGAACAA ACCGTCAGGG
	251 AAAAGCGGCG GAAGAACAGAT CGCCGAACCG TTAAAAATAC ^A ACCGCTAAAAA
	301 CGCTCTAAAG AAACAGAGAGA AAAAGCTTC ^A AAAGAGAGA AAAAGCCGGC
	351 GAAGGAAAGG GTTGACCCCCA AACCAACCC ^A GGAAACAAATC CTCAACAGGG
	401 GCA ^C TCCGA AAAAGCCGGC AgTCCGGCC ^C CCAAAGAAAGT ^T CGGAAAAATG
	451 AA AACGTC ^A GACAAAGCGGG AAGC ATACG ATTATCTGCA ATATGGCCGG
	501 TATGCCGACC GTCAAGGAGGCC GGAAGGGCAG CGTGGCAACAA ^T TGCAATCTT
	551 GGGCATATCT TCCAAGGTGG TCCTGGTATA ^C GGGGGACAT AAAACGCTT
	601 ACCGGGTGCA AAGCGGCAAT ATGCTGCCG ATGCGGTGA

This corresponds to the amino acid sequence <SEQ ID 382; ORF65>:

60	1..ILKPHNQLKE D ^I QFPDPADQN ALSEPDA ^T AEQSDAENAA DKQPVADKAD
	51 EVEEKAGEFE REEPFDGQA ^V R KKLTEEREQ TVREKAQKKD AETVKIQAVK

101 PSKETTEKKAS KEEKKAKEK VAPKPTPEQI LNSGSIEXAR SAAAKEVQKM
 101 XNVRQGGSSXR IIICKWARMPT VRARKGSVPN WQSWAYLPRW SVIRDIKRF
 201 TGCKAAICLP MR*

Further work revealed the complete nucleotide sequence <SEQ ID 383>:

5 1 ATGTTTATGA ACAAATTTC CCAATCCCGA AAAGGTCTGT CCGGTTTTTT
 51 CTTCGGTTTG ATACTGGCGA CGGTCTATTAT TGCCGGTATT TTGTTTTATC
 101 TGACCCAGAG CGGTCAAAT GCGTTCAAAA TCCCGGCTTC GTCGAAGCAG
 151 CCTGCAGAAA CGGAAATCTT GAAACCGAAA AACCAAGCTTA AGGAAGACAT
 201 CCAACCTGAA CGGCCGATC AAAACCCCTT GTCCGACCGG GATGCTGGCA
 251 CAGGGCAGGAG CGACTCGGTG CGGGAAAAG CTGGCACACA GCAGCCGGTT
 301 CGCGATTAAG CGCGCGGGT TGPMGAARRA CGGGCGGAGC CGGAACGGGA
 351 AGAGCGGAC GGACAGGGCG TGCGTARGNA AGCGCTGACG GAAGAGCGTG
 401 AACAAACCGT CAGGGGAAAAGG CGCCAGAAA AGAGTGCAGA AACGGTTAAA
 451 AAAAACCGGG TAAACCGCT TAAGAACAAA GAGAAAAAAGG CTTCAAAAGA
 501 AGAGAAAAGG GGGCGGAAGG AAAAGATTCG ACCAAAACCA ACCCGGGAGAC
 551 AAATCTCAAG CAGCGGGAC ATCGAAAACAG CGCGCAGTGC CGCGCOCAAA
 601 GAATGCGAA AGATGAAAN GTCCGACMA CGGGAAAGCAA CGCATTTATCT
 651 GCAATGGG GCGTAGTCCG ACCGTCAAG CGCGGAAAGGG CAGCGTGC
 701 AACTGGCAAT CTTGGGATA TCTTCGGANG TGTCGGTTA TCAGCGGGGA
 751 CATAAAACGC TTTACGGCG CGAACAGGC AATATGCTG CGCGATGCGGT
 801 GAAAAGATC CAGGGCAGGT TGAAAAAAAC TGAAGTCGCC AGCGCTGATCC
 851 GTTCTATCGA AAGCAAATAA

This corresponds to the amino acid sequence <SEQ ID 384; ORF65-1>:

25 1 MFWNKFSQSG KLGSGFFFGI ILATLVIAGI LEYLNQSGQN AEKIPASSKQ
 51 PAETEILPKP NQPKEDIQPE PADQNALSEP DAAATEEQSD AEKAADQKPV
 101 ADKADEVEEK AGEPEQQEPD GQAVRKALQT EEREQTVREK AQKKDAETVK
 151 KQAVPKSKET EKKASKEEKK AAKEVAPPK TFEQINLSGS IEKARSAAAK
 201 EVQKMTSDE AEATHYLQMG YAADRQSAEC QRALKLAIIGI SSKVVGYQAG
 251 HTLYLVQSG NMSADAVKRM QDELKKHVEA SLIRSTIESK*

30 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N. meningitidis* (strain A)

ORF65 shows 92.0% identity over a 150aa overlap with an ORF (ORF65a) from strain A of *N. meningitidis*:

			10	20	30
35	orf65.pep		ILKPHNQLKEDIQPDADQNALSEPAADE		
	orf65a	IIAGILFYLNQSGQNAFKIPVPSKQPAETEILKFKNQPKREDIQPFADQNALSEPAADE	30	40	50
			60	70	80
40	orf65.pep	40 50 60 70 80 90			
	orf65a	AEQSDAENAADQPVADKADAEVEEKAGEPEREEPDGQAVRKALTEEREQTVREKAAKKD	90	100	110
			120	130	140
45	orf65.pep	100 110 120 130 140 150			
	orf65a	AETVKIQAVKPKSKETEKKASKEEKKAAKEKVKAPKPTFEQILNSGSIEKARAAKEVQKM	150	160	170
			180	190	200
50	orf65.pep	160 170 180 190 200 210			
	orf65a	AETVKQAVKPKSKETEKKASKEEKKAAKEKVKAPKPTFEQILNSGSIEKARAAKEVQKM	210	220	230
			240	250	260
55	orf65.pep	XNVRQGGSSXR IIICKWARMPTVRARKGSVPN WQSWAYLPRW SVIRDIKRTGCKAAICLP			
	orf65a	KTFDKAETHYLQMGAYADRRSAEGQRALKLAIIGLISSSKVVGYQAGHKTLYRVQSGNNMSAD			

The complete length ORF65a nucleotide sequence <SEQ ID 385> is:

1 ATGTTTATGA ACAAATTTC CCAATCCCGA AAAGGTCTGT CCGGTTTTTT
 51 CTTCGGTTTG ATACTGGCGA CGGTCTATTAT TGCCGGTATT TTGTTTTATC

5	101	TGAAAGCAGAG	CGGTGTAAAAG	GCGTTGCAAAA	TCCGGGTTTC	TGGAGACGAC	
	151	CTCGGAGAAG	CGGAAATCTC	GAACCGAAAAA	AACCGAGCTTA	AGGAGACAGAT	
	201	CCAACTGTAA	CGGCCGAGAAG	AAACCGCTT	CTTGCGGACG	AGTGTGCTGC	
	251	AAAGGAGGAA	CGACTGGCAT	CGCGGAAAG	CTGGCAGAAC	CGACGGCGTGT	
	301	GGCGCACAAAG	CGCGAGGATG	TGAGGAAAGG	AGCGGAGCAGC	CGAGGAGGAGC	
	351	AAAGTCGGAC	GGACAGGCG	CGCGCAAGAA	AAGCTAGC	GAAGAGGCGT	
	401	AAACAAACCGT	CGGGGAAAGA	GCGCAAGGAA	AAGATGCTCGC	AAGCTTGAAG	
	451	AAACAAAGGG	TAAAAGGATC	TAAAGAACAA	GAAGAAAAAA	CTTAAAGAAC	
10	501	AGGAGRAAAAG	GGGGAGAAGG	AAAAGATTCG	ACCCCAAACCG	ACCGGGAAACG	
	551	AAATCTCTAA	CAGCGGCAGC	ATCGGAAAG	CSCCGAGTC	CGCTGCCCCAA	
	601	GAAGTGCAGA	AAAATGGAAAG	GCGCGAAAGCA	CGCGATTAATCT	CGCGATTAATCT	
	651	CGAAATTGGC	GGCTGGATGCC	ACCGCGCGAG	CGCGGAAGGG	CGCGCTGGGG	
	701	AACTTGCCAT	CTTGGCCCAT	TCTCGGAAAGG	FTGGCTGGTTA	TCAGCGGCGG	
	751	CTAAGAACCG	TTTACCGGGT	CGAAAGCGGC	ATAATGTCGTC	CGAGATGGGT	
	801	GGAAAGAAATG	CGGAGGACCG	TGAAAAAAAC	TGAAGTCGCC	AGCTGTGATCC	
	851	GTTCATGCA	AAAGGAAATA				

This encodes a protein having amino acid sequence <SEQ ID 386>

1 MFNMFSQSG KGLSGFPPGL ILATVILAGF LFYLNGNSQN AFKIPVPSK
20 51 PASTELPKPQ NQPEKDFIOPQE PADQNALEED AAEKAQESD NQKADQPV
101 101 ADKADEPKK DAEPEPERKS GQAVRKKKAIT EREBQTVGKEK QAKDAETVK
151 KQAVPKSKET EKKASKEEKK AEKEVKVAPK TEPIULNSLGS IKARASAAK
201 EVOKMKTPD AEATHYQMGK AYADRRAEKG QRAKLKHEA SSSKVGVYQAG
251 HTFLYRVQSO NMSADDAVKM KNLCKHEVA SLIRIESK*

ORF65a and ORF65-1 show 96.5% identity in 289 aa overlap.

		10	20	30	40	50	60
25	orf65a.pep	MFMNKFSQSGKGLSGFFGLILATVIIAGILFYLNQSQNAFKIFPVPSKQPAETEILKPK					
	orf65-1	MFMNKFSQSGKGLSGFFGLILATVIIAGILFYLNQSQNAFKIPASSKQPAETEILKPK	10	20	30	40	50
30	orf65a.pep		70	80	90	100	110
	orf65-1	NQPKDEDIQPEPADQNALSPEDAAKEAQSDAEEKAADQPKVADKADEVEEKADEPEREKS	70	80	90	100	110
35	orf65a.pep			70	80	90	100
	orf65-1	NQPKDEDIQPEPADQNALSPEDAATEQSDAEEKAADQPKVADKADEVEEKAGEPEREPP	70	80	90	100	110
40	orf65a.pep		130	140	150	160	170
	orf65-1	GQAVRKKALTEEREOTVGEAQKQDKAETVKKQAVKPSKETEKASKEKKAAKEVKVAPKP	130	140	150	160	170
45	orf65a.pep			190	200	210	220
	orf65-1	TPEQIQLNSGIEKARSAAAKEVQMKMTPDXRAEATHYLQMGAAYDRRSQAEQGRQAKLAIIGT	190	200	210	220	230
50	orf65a.pep				190	200	210
	orf65-1	TPEQIQLNSGIEKARSAAAKEVQMKMTSOKAATATHYLQMGAAYDRRSQAEQGRQAKLAIIGT	190	200	210	220	230
	orf65a.pep		250	260	270	280	290
	orf65-1	SSKVVGQYAGHKTLYRVQSGNMSADAVKKMQDELKKHEVASLIRSIESKX	250	260	270	280	290

55 Homology with a predicted ORF from *N.gonorrhoeae*

ORF65 shows 89.6% identity over a 212aa overlap with a predicted ORF (ORF65.ng) from *N. gonorrhoeae*:

	30	40	50	60	70	80
ORF65ng	IIAGILLYLNQQGNKFAPSKQPAETEILKLKNQPKFEDIQPEPADQNALSEPVDVAK					
ORF65		: : : :				
			ILKPHNQLKEDIQPPDADQNALSEPDA			
				10	20	30

-245-

		90	100	110	120	130	140	
	ORF65ng	AEQSDAERAADKQPVDADEVEEKAGEPEREPERPDGQAVRKALKTEEREQTVREKAQKKD						
5	ORF65	AEQSDAERAADKQPVDADEVEEKAGEPEREPERPDGQAVRKALKTEEREQTVREKAQKKD	40	50	60	70	80	90
	ORF65ng	150 160 170 180 190 200						
10	ORF65	AETVKIGAVKPSKETEKKAKEEKVAPKPTPEQILNSRSIEKARSAAAKEVQKM	100	110	120	130	140	150
	ORF65ng	210 220 230 240 250 260						
15	ORF65	KNFGQQGSQRIICKWARMNPNGARKGSVPNWQSWAYLPKWSAIRDIKRTFTACKAAICPP	160	170	180	190	200	210
	ORF65	XNVRQGGSXRIICKWARMPTVRAKGSPVNWQSWAYLPKWSVIRRDIKRTFTGCKAAICLP						
20	ORF65ng	MR						
	ORF65							
	ORF65	MR						

An ORF65ng nucleotide sequence <SEQ ID 387> was predicted to encode a protein having amino acid sequence <SEQ ID 388>:

25	1	MFMNKFSQSG KGLSGFFFGL ILATVILIAGI LLYLNQGGQN AFKIPAPSQKQ
	51	PAETEILKLK NQPKEDIQPE PADQNALSEP DVAKRAEQSDF AEKAADKQPV
	101	ADKADEVEEK AGEPEREPDQOAVRKALKTEEREQTVREKAQKKD
	151	KKAVKPSKET EKKASKEEKK AAKEEVAPKPTPEQILNSRS IEKARSAAAK
	201	EVQMKMNFQG GGSQRICKW ARMPNGARK GSVPNWQSWA YLPKWSAIRR
	251	DIKRTFTACKA AICPPM*
30	After further analysis, the complete gonococcal DNA sequence <SEQ ID 389> was found to be:	
	1	ATGTGTTATGA ACAAACTTTC CAAATCGGAA AAAAGGTCTGT CGGGTTCTCTT
	51	CTTGGTTTCA ATATCGGCAA CGGTCAATTAT TTGGCTTATTC
35	101	TGAACCAAGG CGCTGCAAAAT GCGTCAAAAAA TCCCGGGCTCC GTCGGAAGCAG
	151	CTTCAGAAA CGGAAATCTT GAAACTGAAA AACCGCCTTA AGGRAGACAT
	201	CCACCTGGA CGCGGCCGATC AAAACGCTCT GTGCCAGAACGG GATGTTGCGA
	251	AAGRGGCAGA CGACCTGGAT GGGGAAAAAG CTGCCAGCAA GAAGCCCGTT
	301	GCGACCAAGa cccggacgggt TGAAGAAAag GcgGcgAgc cggAGCggAg
	351	AGACCGGAGC ggACGGGAGC TGCCAGAAGA AGCAGCTGact gAGAGcGTG
40	401	AACAACCGt cggggAAAAA GCGGAGaaaga AAGATGCCGA AACGgtTTAA
	451	AAacAAcctg tAAacccgtc tAAAGGAAACCA gggaaaaaaaat cttcaaaaaaa
	501	agaaaaaaaag gcgccggaaa agg aAGTgtc acccaaaaccy accccggaaaC
	551	aaatccctca cccggcggCact atccggaaaaag cgccgtatgtc cgctggccaaa
	601	gaAtgtcaGA AAAtggaaaat ttTggccca gggcGagcc aacgcattAT
45	651	CTGGAAatgtg gcccgtatgc ccggccgtcc gggccggggaa gggcagcgtg
	701	cccaAAatCtgg aAtcttgGgc atatccfccc aagtgtcgG CTATCAGGG
	751	GGACATAAAA CGCTTAAACCG CGTGTCAAAAGc GGCAtatatgt ccggccgatgc
	801	gGTGAAAAAA ATGACGGACO AGTTGAAAAA GCATGGGGttt gcCAGCCTGA
	851	851 TCCGTGcgt TGAAGGCAAA TAA

This encodes the following amino acid sequence <SEQ ID 390>:

50	1	MFMNKFSQSG KGLSGFFFGL ILATVILIAGI LLYLNQGGQN AFKIPAPSQKQ
	51	PAETEILKLK NQPKEDIQPE PADQNALSEP DVAKRAEQSDF AEKAADKQPV
	101	ADKADEVEEK AGEPEREPDQOAVRKALKTEEREQTVREKAQKKD
	151	KQAVKPSKET EKKASKEEKK AAKEEVAPKPTPEQILNSRS IEKARSAAAK
	201	EVQMKMNFQG GGSQRICKW ARMPVRSAAE GQRALKALILG ISSEVVGYQA
55	251	GHKTLYRVQS GNNSADAVKKE QDELKLKHGV ASLIRAIEGH *

ORF65ng-1 and ORF65-1 show 89.0% identity in 290 aa overlap:

		10	20	30	40	50	60
60	orf65-1.pep	MFMNKFSQSGKGLSGFFFGLILATVILIAGILFYLNQSQNAFKIPASSKQPAETEILKLK					
	orf65ng-1	MFMNKFSQSGKGLSGFFFGLILATVILIAGILFYLNQSQNAFKIPASSKQPAETEILKLK	10	20	30	40	50

		70	80	90	100	110	120
5	orf65-1.pep	NQPKEQDIPPEPADQNALSPEDAATEAQSDAEKAADKQPVADKADVEEKKAGEPEREED					
	orf65ng-1	NQPKEQDIPPEPADQNALSPEDVAKAEAQSDAEKAADKQPVADKADVEEKKAGEPEREED	70	80	90	100	110
10	orf65-1.pep		130	140	150	160	170
	orf65ng-1	GOAVRKALKALTEEREQTVERKQAKKDAETVKQKAVPKSKETEKKAKESKKAKEKVKAPK	130	140	150	160	170
15	orf65-1.pep		190	200	210	220	230
	orf65ng-1	TPEQILNGSIEKARSAAAKEVQKMNTSDKAEEATHYL-QMGAYADRQSAAEGQRAKLIALTG	190	200	210	220	230
20	orf65-1.pep		240	250	260	270	280
	orf65ng-1	ISSKVVGYYQAGHKTLYRVQSGNMSADAVKRMQDELKKHEVASLIRSIESKX	250	260	270	280	290

- 25 On this basis, including the presence of a putative transmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 46

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID

- 30 391>

35	1	ATGAAACCAACC	ACTAACCTT	CCTCCACCCCG	TTCCTACGG	GTKTCCTTCCG
	51	CGGAAACGAC	TGCTACCGTA	TTGGCCGGGA	ATTAAAGCAGG	GCGTTTGGS, s,
	101	TCCAACTCC	CCCCGATCAT	AACCCGTTTT	GCGCTATCTC	GGCTCTTAA
	151	ACAGGACCGG	TAAGCGACGA	TACGCCAATC	GGCGGTACAT	TCGGATTAAAT
	201	CGGACGATGC	GGCCCTTTC	TGCGAACAAAC	GGCGCTCTGG	GAQAAATTAT
	251	TATACACCGC	GGCCAAACCTC	CTGCTGCTCT	TTTAAAGCTT	ATACTTGAGCC
	301	GTTGATTTCTT	CTTGGCCGG	AAAATCCGGG	AAAATCGGCCA	GGGAACTATG
	351	GGCGAACCTTC	AAACCCGATAC	TCACCAACCT	TTTACCATCA	AAACCTCATAC
40	401	CGCGCTCTGCCT	L6GCGCTGG	ATTTATGGG	GTCGTCGTCG	GGGGGACTGAT
	451	GGTTTACAGGG	GCTGCTTTA	CGCGCTGGGA	AgGGCTAGTG	CGGAAACAGGGG
	501	CGGGTTTATAT	ATGCTGCTTG	TTCGACCTGGG	TAGCCGTCGCC	AACTCTT+TAG
	551	CAATCGGCAT	TTT+TCCCG	CAACTGAAWA	AAATCATCGA	AAACCCGAT
	601	ATCGCGCTTG	GTACGGGAT	ATCGCTATCA	TTATGGGCAT	TATGGAAACT
	651	TGGCCGCTTG	TGGCTGTAA			

- 45 This corresponds to the amino acid sequence <SEQ ID 392; ORF103>;

1 MNNDITFLTL FLLGKFGGTH CIGMCGGLSS AFXXOLPHI NFRWFLLLNL
 51 TGRVSYTTA KLGKFLGLQV GVSLDQTWRP QNLYTAANL NLLFLGLGLYS
 101 GISSLAARIK GLKGPIRWNL NRRNLLPLPI KSIKAPCIALVG ILWGWLPIL
 151 VYASVALO GSAGATAGLVY MAFALGTPLP NLLAIGIFSL QLKXIMONRY
 201 RILCTGLSVA LNALWKLVWL WI*

Further work elaborated the DNA sequence <SEQ ID 393> as:

1 ATGAAACCGC ATGACATCT CCT AACCCCGT TTCTCTGG GTTCTTCGCG
 51 CGGAAGGCAC TGGCATCTGA TTGTGGGGGG ATTAAAGCAGC GGTTTCTGGCG
 101 TCCAACCTCC CCCGCATATC AACCGCTTTT GGCTGTATCT GCTGCTTAAC
 151 ACAGACGGG TAGAACGGT TAGGGCAATC GGCGCTTAGC TGGATTATAA
 201 CGGAGACGGT GGCGTTTCA CGGAAACAC CGGGCTGCTC CAGAATTTT
 251 TATACCGGC CGCCAAACTC CTGCGCTCTT TTAGGCTT ACTATGCGC
 301 GGTTATTCCT CTTGGGGC AAAATCGAG AAAATGCCG CAGGATATG

351 GOGGAACTGG AACCAGATAC TCAACCCGGT GTTACCCATA AAATCCATAC
 401 CGCGCTGCCT TGCGGTGGGA ATATTATGGG GCTGGCTGCC GTGCGGACTG
 451 GTTTTACAGGG CGTCGCTTTA CGGCGTGGGA AGCGGTAGTG CGCGAACGGG
 501 CGGGTTATAT ATGCTTGCTT TTGCACTGGG TAGCGTGCCTT AATCTTTAG
 551 CAATCGCAT TTTTCTCTG CAACTGAAA AAATCATGCA AAACCGATAT
 601 ATCCGGCTGT GTACGGGATT ATCCGTATCA TTATGGCAT TATCGAAACT
 651 TGGCGCTCTG TGGCTGTAA

This corresponds to the amino acid sequence <SEQ ID 394; ORF103-1>:

1 MNHDITFLTL FLLGGFFGTH CIGMCGGLSS AFALQLPPHI NRFWLILLLN
 10 51 TGRVSSYTAI GLILGLIGOV GSLSLDOTRVL QNLYTAANL LLLFLGLYL
 101 GISSLAAKIE KIGKPIWRNL NPILNRLLPI KSIPACLAVG IIWOWLPGL
 151 VYSASLYALG SGSAATGGLY MLAFALGTLF NLLAIGIFSL QLKKimQNR
 201 201 IRLCTGLSVS LNWLWKLAVL WL*

Computer analysis of this amino acid sequence gave the following results:

15 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF103 shows 93.8% identity over a 222aa overlap with an ORF (ORF103a) from strain A of *N. meningitidis*:

	orf103-pep	10	20	30	40	50	60	
20	orf103a	MNHDTITFLTLFLGXFGGTHCIGMCGGLSS AFALQLPPHI NRFWLILLN						
	orf103a	MNHDITFLTLFLLGFFGTHCIGMCGGLSS AFALQLPPHI NRFWLILLN	10	20	30	40	50	
25	orf103-pep	GLILGLIGQVGVSLSQDTRVLQNLTYTAANL LLLFLGLYLSSLAAKIE KIGKPIWRNL	70	80	90	100	110	120
	orf103a	GLILGLIGQVGVSLSQDTRVKQNLTYTAANL LLLFLGLYLSSLAAKIE KIGKPIWRNL	70	80	90	100	110	120
30	orf103-pep	NPILNRLLPIKSIPACLAVGILWGWLPCGLVY SASLYALGGSAAATGGLYMLAFACTLP	130	140	150	160	170	180
	orf103a	NPILNRLLPIKSIPACLAVGILWGWLPCGLVY SASLYALGGSAAATGGLYMLAFACTLP	130	140	150	160	170	180
35	orf103-pep	NLLAIGIFSOLQXLKIMQNRYIRLCTGLSVSLWALWKLAVLWLX	190	200	210	220		
	orf103a	NLKAIGIFSOLQXLKIMQNRYIRLCTGLSVSLWALWKLAVLWLX	190	200	210	220		

The complete length ORF103a nucleotide sequence <SEQ ID 395> is:

1 ATGAACCANG ACATCACTTT CCTCACCTTG TTCTCTACTCG GTTCTCTCCG
 51 CGGAACGCGAT CGGCGCTGTT TGTCGCTGGTA ATTAAAGCAGC CGCTTGTGCG
 101 TCCAATCTCC CCCGCATATC AACCGCTTNT GGCTGATCT CTGCTCTTAAC
 151 ACAGCAAGGG CGTCGCTTCA CGACGCAATC GGCGCTGATAC TCGGATTTAT
 201 CGGCAAGCTG CGGCCTTCTAC TGACCAAAAC CGGCCTCTNC CAGAAATTTT
 251 TATACACGGC CGCCAACCTC CTGCTGCTCTT TTTCAGGCTT ATACTTGAGC
 301 GGTTATTTCTT CCTTGGCGG AAAATCGAG AAAATCGGA AACCGATATG
 351 CGCGAACTCTT AACCCTGATAC TCAACCGGGT GTTACCCATA AAATCCATAC
 401 CGCGCTGCCT TGCGGTGCGA ATATATGGG GCCTGGCTGCC GTGGGAGCTA
 451 GTTTACAGGG CGTCGCTTCA CGGCGTGGGA AGCGGTAGTG CGCGAACGGC
 501 CGGGTTATAT ATGCTTGCTT TTGCACTGGG TAGCGTGCCTT AATCTTTNGC
 551 CAATCGCAT TTTTCTCTG CAACTGAAA AAATCATGCA AAACCGATAT
 601 ATCCGGCTGT GTACGGGATT ATCCGTATCA TTATGGCAT TATCGAAACT
 651 TGGCGCTCTG TGCGCTGTA

This encodes a protein having amino acid sequence <SEQ ID 396>:

1 MNXDITFLTL FLLGGFFGTH CIGMCGGLSS AFALQLPPHI NRFWLILLN
 51 TGRVSSYTAI GLILGLIGOV GSLSLDOTRVL QNLYTAANL LLLFLGLYL
 101 GISSLAAKIE KIGKPIWRNL NPILNRLLPI KSIPACLAVG IIWOWLPGL

-248-

151 VYSASLYALG SGSAATGGLY M~~L~~A~~F~~ALGTL~~P~~ NLXAI~~G~~IFSL QLXXKIMQNRY
 201 I~~R~~LC~~T~~GLS~~V~~S LW~~A~~WLK~~L~~AVL WL*

ORF103a and ORF103-1 show 97.7% identity in 222 aa overlap:

		10	20	30	40	50	60
5	orf103a.pep	MNXDITFLTLFLLGFFGGTHCIGMCGGLSSAFALQLPPHINRXLWLLLNTRGVSSYTAI					
	orf103-1	MNH D ITFLTLFLLGFFGGTHCIGMCGGLSSAFALQLPPHINR F WLWLLLNTRGVSSYTAI	10	20	30	40	50
10	orf103a.pep	GLILGLIGQVGVSLSLDQ T RVXQN I LYTA N LLFLGLYLSGSIS L AAKIEKIGKP W RNL	70	80	90	100	110
	orf103-1	GLILGLIGQVGVSLSLDQ T RV L QN I LYTA N LLFLGLYLSGSIS L AAKIEKIGKP W RNL	70	80	90	100	110
15	orf103a.pep	NPILNRLLP I KSI P ACLA V GIL W GLPC G LVY S AS L YALGSGSAATGGLYMLA F ALGTL P	130	140	150	160	170
	orf103-1	NPILNRLLP I KSI P ACLA V GIL W GLPC G LVY S AS L YALGSGSAATGGLYMLA F ALGTL P	130	140	150	160	170
20	orf103a.pep	NLXAI G IFSLQLXKIMQNRYI R LCT G LS V SLW A WLK L VLWLX	190	200	210	220	
	orf103-1	NL A AI G IFSLQLKKIMQNRYI R LCT G LS V SLW A WLK L VLWLX	190	200	210	220	
25	orf103a.pep						
	orf103-1						

Homology with a predicted ORF from *N.gonorrhoeae*

ORF103 shows 95.5% identity over a 222aa overlap with a predicted ORF (ORF103.ng) from *N.*

30 *gonorrhoeae*:

	orf103.pep	MNH D ITFLTLFLLGFFGGTHCIGMCGGLSSAF X QLPPHINR F WLWLLLNTRGVSSYTAI	60
	orf103ng	MNH D ITFLTLFLLGFFGGTHCIGMCGGLSSAFALQLPPHINR F WLWLLLNTRGVSSYTAI	60
35	orf103.pep	GLILGLIGQVGVSLSLDQ T RV L QN I LYTA N LLFLGLYLSGSIS L AAKIEKIGKP W RNL	120
	orf103ng	GLILGLIGQ L IQ L ISLDQ T RV L QN I LYTA N LLFLGLYLSGSIS L AAKIEKIGKP W RNL	120
40	orf103.pep	NPILNRLLP I KSI P ACLA V GIL W GLPC G LVY S AS L YALGSGSAATGGLYMLA F ALGTL P	180
	orf103ng	NPILNRLLP I KSI P ACLA V GIL W GLPC G LVY S AS L YALGSGSAATGGLYMLA F ALGTL P	180
45	orf103.pep	NL A AI G IFSLQLKKIMQNRYI R LCT G LS V SLW A WLK L VLWL 222	
	orf103ng	NL A AI G IFSLQLKKIMQNRYI R LCT G LS V SLW A WLK L VLWL 222	

The complete length ORF103ng nucleotide sequence <SEQ ID 397> is:

1	ATGAA C CTAC	ATC A TC T CTC	CCTCACC T TT	TCCTGCT C	GT T TCTT C CGG
51	CGGAA C TCAC	TG C AT G GT T	TG T CGCCGCC	ATTAAGC A G	GGGTT T GGC
101	TCC A CT C CC	CC G GC A T T TC	AAC G CGCT T TT	GGCTGATT T CT	GT G TCTTAAC
151	ACAGGAC G GA	TAAGCAG T TA	TAC G GGCA A TC	GGC C CT G AT G C	T G CGGATT A AT
201	CGGACAA C TC	GGCATTTC C AC	TG G AC G AA A AC	ccgcgcTCTG	CAA A AT T TT
251	Latacaca g c	etccca C CTC	CTG T GT G CT T	TTT T AG G CT T	AT A CTTGAGC
301	GGTATT T CTC	CT T TGGCGCG	AAAAT T CG G A	AAAAT T CG G A	AACCGA T AT G
351	GGCGCAAC T CT	AAC C CG A T A C	TCAAC C GG G T	GCT G CCCAT A	AAAT C AT C AT A
401	CCGGCT G CT	TG T CT G CG G	AT A TT T AT G GG	GCT G GC T CG G	GT G CGGACT G
451	GT T TA C AG G	CAT A CT T TA	CGG C CT G GG A	AGCGG T ACT G	CGACAAC C GG
501	CGGACT G TA T	AT G CG T GC T	TT G CA T GG G	TAC G GT G CC C	AA T CT T TT T GG
551	CA T CG G CA T	TTT T CC C CT G	CA A CT G AAA A	AA A TC A GT C A	AA C CG G AT T
601	AT C CG C CT G	GT A CA G GG T	AT C CG T AT C CA	TT A T G GG C AT	TAT G GAAGC T
651	TG C CG T CT G	TG G CG T GT A			

This encodes a protein having amino acid sequence <SEQ ID 398>:

1	MNHADITFLTL	FLLGGFFGTH	CIGMCGGLSS	<u>AFA</u> LQLPPHI	NRFWLILLLN
51	TGRISYYTAI	GLMLGLQQL	GISLDOTRVL	QNLITYTASN	LLLFLGLYLS
101	GISSLAAKIE	KICKPIWRLN	NPLINRRLPI	KSPACLAvg	ILWGWLPCL
151	VYASLYALG	SGSATTGGGLY	MLAFALGTLP	NLLAIGIESL	QLKKIMONRY
201	IRLCTGLSVS	LWALWKLAVL	WL*		

In addition, ORF103ng and ORF103-1 show 97.3% identity in 222 aa overlap:

	orf103-1.pep	10	20	30	40	50	60
10	orf103ng	MNHADITFLTL	FLLGGFFGTH	CIGMCGGLSS	<u>AFA</u> LQLPPHI	NRFWLILLLN	GRISSYTAI
		10	20	30	40	50	60
15	orf103-1.pep	70	80	90	100	110	120
15	orf103ng	GLILGLIGQGV	VSLS	DOTRVL	QNLYTAPAN	LLFLGLYLS	GISSLAAKIE
		70	80	90	100	110	120
20	orf103-1.pep	130	140	150	160	170	180
20	orf103ng	NPILNRLLP	IKPSI	IPACLA	VGIL	WGWLPCGLV	YSASLYALGS
		130	140	150	160	170	180
25	orf103-1.pep	190	200	210	220		
25	orf103ng	NLLAIGIFS	LQLKKIM	QNYI	YIRLCTGLS	VSLWALWKLAVL	WLNX
		190	200	210	220		

30 Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 47

35 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 399>:

1	ATGGAAAACC	AAAGGCCGT	CCTAGGCTT	CGCTTGGCAC	TTTGGCGGC
51	GATGACGTGG	GGAAAGCGTC	CGAT...TCGT	CGCGCAGGTTA	TTGAAGTTTG
101	TCGAGTCGCC	GAACGCTGGT	TGGGTGCGGT	TTACCGCTGC	GGCGCCGGTA
151	TTGGTTGTTT	TGCTGGCACT	GGGGCGCGC	CTGGCGAACG	GGCGAGATT
201	TTGGTTGTTG	CTGCTTCAAG	TCGCTGCTGC	TCGGCGTGGC	GGCGATTTCG
251	GCACAACTTG	TGGTGTATTG	CGAACGGCTG	CATTATATT	CGCGGACAC
301	GACCCAGGT	TTGGGGCAGA	TTTCGGCOST	TACGATGATT	GTWGTCGGTG
351	TGTTGGTGT	TAAGACCGG	ATGACTGGCG	CTCAGAAAAT	CGCGTTGGTT
401	TTGCTGCTTG	CGGGTTGCT	TATGTTATT	AAACGATAAAAT	TCGGCGAGTT
451	GTGGGGTTG	GGGGGGTATG	C. AAAGGGCG	GTGCTGTGTT	GGGGCAGGCA
501	GTATGGCATG	GGTGTGTAAT	GGCGTGGCG	AAAAGCTGCT	CTCGCGGCAA
551	TTGGGGCGCG	AAACAGATTCT	GCTGTTGAT	TATGCGGCAA	GTGCGCCCGT
601	GTTCTCTGCC	TTTGGCGGAC	CGGAAACACAT	CGGAAATGAT	GAACGGTACGT
651	TGGCGTGGGT	ATGTTATGCG	TATGCTGCT	TGAATACGTT	AAATCGGTTAC
701	GGCTCTTGG	GGGAGGGCGT	GAACATTTG	GAGGCTCCA	AAAGTCAGCGC
751	GGTAACAAAC	TTGTCCTCCC	TGTTAACCGT	AAATAATACT	TTGCTCGGGC
801	ATTATGTGAT	GCCTGAAACT	TTTGGCGGCG	CGGA..	

This corresponds to the amino acid sequence <SEQ ID 400; ORF104>:

1	MENORPLLLGF	RLLAALAMTW	GTLPLXSVROV	LKFVDAPTLV	WVRFTVAAG
51	LFVLLALGGR	LPKRKRDFSWK	SFRLLLVLGA	GISANFVLIA	QGLHYISPTT
101	TQVILWQISPF	TMIVVGVLF	KDRMTAAOKI	GLVLLLAGLI	MYFNDFRGEL
151	SGLGAYXKGV	LLCAAGSMAW	VCNAVAQKLII	SAQFGPMQOIL	LLIYAASAAV
201	FLPFEEPAHII	GSMDGTLAWV	CIAYCCLNTL	IGYGSFGEAL	KHWEASKVSA

-250-

251 VTTLLPVFTV INTLLGHYVM PETFAAP...

Further work revealed further partial DNA sequence <SEQ ID 401>:

1 ATGGAAACCC AAAAGGCCCT CCTGGATTCCTG CGCTTGGCAC TTTCATGGCCT
 5 GATGACCTGG GGAAGCCCTG CGATTCGGCG GGCAGGAGTT
 10 TCGAGGGCCC GAGCCCTGG TGCGGCGCTT TAACCTGGCG GGCAGGCGTA
 15 TTGGTTGTTT TGCTGGACCT GGGCGCCGG CGCCGAAAG
 20 TTCTTGCGNC TCAATTCGGC TGCTGGCTGT CGCGGCTGCC
 25 CAACATTGTTG CTGCAATGGC AAAGGGCTTC ATTATAATTTC
 30 ACGCAGGTTT TGTCGGCATG ATTCGGCTTT AGCATGATTCG TTGCTGGTCT
 35 GTTGTGTTT AAAGACCGGA TGACTGGCC CGAAGAACCTG
 40 TCTGCTTGG CGGGTTGCTT ATGTTTTTA ACCTGATAAT CGCGCAGTTG
 45 TCGGGTTGG CGGGCTATGG GAAAGGGCGT TGCTGCTG CGCCAGGCAAT
 50 TATGGCATGG GTGTYITATG CGCTGGCGCA AAAGCTGTC TGCGCAGGAT
 55 TCGGGCCCGA ACAGATTTG CTGTTGATTT ATTCGGCAAG TCCGCGCTG
 60 TTCTGGCTT TTGGCCAACT GGACACATTC GGAAGTTTGG ACCTGATTTG
 65 GGGCTGGTTT TTGTTGGCTT ATTCGGCTT GAATACGTTA ATCGGCTG
 70 GCTCTGGTCC GAGGGCGT AAAACATGGG AGCTCTTCAA AGTCAGGCC
 75 GTCAAACCTG TTCTGCCCGT GTTACCGTA ATAATWTCCTG TGCTCGGGCA
 80 TTATGTGATG CTCGAAACCT TTGCGGCCG GGA...

20 This corresponds to the amino acid sequence <SEQ ID 402; ORF104-1>.

1 MENQRPLLGF ALALLAAMTW GTLPIAVRQV LKFVDAPTLW WVRFRTVAAAV
51 LEFLVLALRGH LPKRGRDFSWC SFRLLNLGKQI GIASVANFLY QGLHYISPT
51 TQVLVQISPF TMVVGGVLFV KORMTAQOKI GLVLLNLGIL MFNFNDKGEL
51 SGGLAYAKGV LLCAAGSMAW VCYAQAKLII SAQFGPQQIL LLIYAASAAB
1 FLPFPAEPAH GSDLGDTLAWV CFAYCCTNLG IGYGSFGEA LKHWEASKVSA
51 VTTTFLVFVTV IXKLGHYQFPTT FEFAP...
51

Computer analysis of this amino acid sequence gave the following results

Homology with hypothetical HJ0878 protein of *H. influenzae* (accession number U32769)

QBEL04 and HJ0878 show 40% aa identity in 277aa overlap:

30	orf104	4	QRLPLGFRLALLAAMTWTGTLPKSVRQLVKFVDAFTLVWXXXXXXXXXXXXXXXP- 62 Q+PLLGF AL- AM WG-LP +++QVL +AA T+VW P	
	H10878	3	QQLLGGTFATIAMAWSPLAIKQLVSVNMAQTIVWYRFIAAVSLLALLAYKKQLE 62	
35	orf104	63	--KKRDFSWCSPLRLLLGVAGVIGSANFVLIAGQLHYISPTTTQVLWQISPFMTIVVGVLVF 120 K R ++W ++LG-V +--+NNFL + L-YI P- + + + S F M++ GVL+F	
	H10878	63	LMLVKQYAW---IMLGVLGTSNLFLSSLSNVLYIEPSVQAFIChLHSFFGICLGVGLF 118	
	orf104	121	KDRMTRAAQKIXXXXXXXXXXMFNDKFPGELSGLGAYXKGVLCAAGSMWAHVCAVRQKLL 180 K+++ QKI ++FND+P +GL Y GV- G++ WV +AQKL+	
40	H10878	119	KEKLGLHQKQFLFLLGLGFLNDFRDFADAGLNQYSTVILGVLGWVYAVGMQAKL 178	
	orf104	181	SACQFPQ011LLIYAASAAVFLPFPFAEPAHIGSMGTLAWCIIAYCCNLTLIGYSFGEAL 240 +F QOILL+Y A F+P A++ + LA +C YCCNLTLIGYS+ EAL	
	H10878	179	179 LRKFNQSNMMLYLGCAIAFMADFSQVQELT- PLALICFCIYCCNLTLIGYSVAEAL 237	
45	orf104	241	KHWEASKVSATVLLPVFTVINTLGHYMPETFAAP 277 +W SKVS V TL+P+T++ + HY P FAAP	
	H10878	238	238 NRBLKQDSKVSVTLLVPLFETIPLSHIAHYFSPDFAAP 274	

50 Homology with a predicted ORF from *N. meningitidis* (strain A)

ORF104 shows 95.3% identity over a 277aa overlap with an ORF (ORF104a) from strain A of *N. meningitidis*.

-251-

	orf104.pep	LPKRRDFSWCSFRLLLLGVAGISANFVILIAQGLHYISPTTQVLWQISPTTMIVVGVLVF
5	orf104a	LPKWRDFSWCSFRLLLLGVAGISANFVILIAQGLHYISPTTQVLWQISPTTMIVVGVLVF 70 80 90 100 110 120
	orf104.pep	130 140 150 160 170 180
10	orf104a	KDRMTAAQKIGLVLLLAGLMLYFNDFKFGELSGLGAYAKGVLLCAAGSMAWWCYAVAQKLL KDRMTAAQKIGLVLLLAGLMLYFNDFKFGELSGLGAYAKGVLLCAAGSMAWWCYAVAQKLL 130 140 150 160 170 180
15	orf104.pep	190 200 210 220 230 240
	orf104a	SAQFPFQQIILLYAASAAVFLPFAELAHIGSLDTLANVCFAYCCLNLTIGYGSFGEAL SAQFPFQQIILLYAASAAVFLPFAELAHIGSLDTLANVCFAYCCLNLTIGYGSFGEAL 190 200 210 220 230 240
20	orf104.pep	250 260 270
	orf104a	KHWEASKVSAVTLLPVFTVINTLLGHYVMPETFAAP KHWEASKVSAVTLLPVFTVINTLLGHYVMPDTFAAFDMNGLYAGALVVGGAVTAAGV 250 260 270 280 290 300

The complete length ORF104a nucleotide sequence <SEQ ID 403> is:

1	ATGAAAACC	AAAGCCCGCT	CCTAGGCTTC	GGCTTGGCAC	TTTGCCGCC
51	GATGACGTGG	GGAACGCGTC	CGATTGCGCT	GGCGCAGGTA	TTGAAGTTTG
101	TCTGTCGGGC	GACGTCGGTC	TGGGTGGCGT	TTACCGTGGC	CCCGCGGCTA
151	TTGTTGGTT	TGCTGGCAT	GGCGGGCGCG	CTGCGGAAGT	CCCGGGATT
201	TTCCTGGTC	TCTTCAGGC	TGCTGTGTC	CGCCCTGGCC	CGCATTTCCG
251	CAAATTCTGT	GCTGATTCGC	CAAGGGCTGC	ATTATATTTTC	CCGCACCCAG
301	ACCCAGGTG	TTCGGCAGAT	TTCCCGCTTG	ACGATGATTG	TTGTCGGGT
351	GTGGGGTTT	AAAGACCGGC	TGACTGCGCC	TCAGAAATTC	GGCTTGGTT
401	TGCTGGCTGC	CGGGTTCCTT	ATGTTTTTA	ACGATAAAATT	CGGGGAGCTG
451	TGGGGTTTG	GCGGCTATCC	GAAGGGCGTG	TTGCTGNTG	CCGAGGCG
501	TATGGCATGC	CTGTTTATG	CGCTGGGGCG	AAAGCTGCTG	TCCGGCCAAAT
551	TCGGGGCCA	ACAGATCTG	CTGTTGATT	ATCGCGCAAG	TCCGGCCGTG
601	TTCCCTGCC	TTGGCGAAC	GGCACACATG	GGAACTTTTG	ACGGTACCGT
651	GGCTGGGGT	TGTTTGGCGT	ATTCGTCGT	GAATACGTTA	ATCCGGTAGC
701	GCTGTTGC	CGAGGGCGTG	AAACATTTGG	AGGGCTTCCAA	AGTCAGCCGG
751	GTAAACACCT	TGCTCCCGT	GTTCACCGT	ATATTTCCTT	TGCTCGGGCA
801	TTATGCTGAT	CTCTGATCTT	GGGGCCCCCG	GGATATGAAC	GGTTTGGGT
851	ATGCCGGCCC	ACTGCTCGT	GTCCGGGGTG	CGCTTACCGC	GGCGGTGGGG
901	GACAGGCTGT	TCAAAACCGG	CTAG		

This encodes a protein having amino acid sequence <SEQ ID 404>:

1	MENQRPLLG	FA1ALLAAMTW	GTLPIAVRQV	LKFVDAPTLV	WVRFTVAAG
51	LFVLLALIGR	LFKWRDFSWC	SFRLLLGVVA	GISANFVILIA	QGLHYISPTT
101	TOVILWISPF	TMIVVGVLYF	KDRMTAAQKIL	GLVLLLAGL	MFNFNDKFGEL
151	SGLGAYAGV	LLCAAGSMW	VCAVAAQKLL	SAQFPFQQIIL	LLIYAASAAV
201	FLPFAELAHI	GSLDGTLLAWV	CFAYCCLNLT	IGYGSFGEAL	KHWEASKVS
251	VTLFLPVFTV	IFSLLGHYV	PDTFAAPDM	GLGYAGALVV	VGGAVTAAGV
301	DRLFKR*				

ORF104a and ORF104-1 show 98.2% identity in 277 aa overlap:

	orf104a.pep	10	20	30	40	50	60
55	orf104-1	MENQRPLLG	FA1ALLAAMTW	GTLPIAVRQV	LKFVDAPTLV	WVRFTVAAG	
		10	20	30	40	50	60
60	orf104a.pep	LPKWRDFSWCS	FRLLLLGVAGISANFVILIA	QGLHYISPTT	QVLWQISPTTMIVVGVLVF		
	orf104-1	LPKRRDFSWCS	FRLLLLGVAGISANFVILIA	QGLHYISPTT	QVLWQISPTTMIVVGVLVF		
		70	80	90	100	110	120
65	orf104a.pep	130	140	150	160	170	180
	orf104-1	KDRMTAAQKIGLV	LLLAGLMLFFNDFK	FGELSGLGAYAK	GVLLCAAGSMW	CYAVAQKLL	

	orf104-1	KDRMTAAQKIGLVLLLAGLMFFNDKGELSGLGRAYAKVLLCAARGSMAWCYVAQKLL	130	140	150	160	170	180
5	orf104a.pep	SAQFGPQQIILLLIYAAASAVALFAEALAHIGSLDGTIAWVCFAYCLNTLIGYGSFGEAL	190	200	210	220	230	240
	orf104-1	SAQFGPQQIILLLIYAAASAVALFAEPAHIGSLDGTIAWVCFAYCLNTLIGYGSFGEAL	190	200	210	220	230	240
10	orf104a.pep	KHWEASKVSATVTLLPVFTVIFSLLGHVYVMPDTFAAFAAEDMNGLGYAGALVVVGAVTAAVG	250	260	270	280	290	300
15	orf104-1	KHWEASKVSATVTLLPVFTVIXXLLGHVYVMPETFAAFAAEDMNGLGYAGALVVVGAVTAAVG	250	260	270			

Homology with a predicted ORF from *N. gonorrhoeae*

ORF104 shows 93.9% identity over a 277aa overlap with a predicted ORF (ORF104.ng) from *N. gonorrhoeae*:

20	orf104.pep	MENQRPLLGFLALLAAMTWGTLPLXSVRQLKFVDAPTLVWWRFTVAAAVLFVLLALGGR	60
	orf104ng	MENQRPLLGFLALLAAMTWGTLPIAVRQLKFVDAPTLVWWRFTVAAAVLFVLLALGGR	60
25	orf104.pep	LPKRDRFSWCFSRFLLLLGVGAGISANFVLAIAQGLHYISPTTQVLWQISFPTMIVVGVLVF	120
	orf104ng	LPKRDRFSWSIFRFLLLLGVTGISAIFVLAIAQGLHYISPTTQVLWQISFPTMIVVGVLVF	120
	orf104.pep	KDRMTAAQKIGLVLLLAGLMMYFNDKFEGELSGLGRAYKGVLCAAAGSMAWVCNRAVKLL	180
30	orf104ng	KDRMTAAQKIGLVLLLGVLMFFNDKGELSGLGRAYKGVLCAAAGSMAWVCNRAVKLL	180
	orf104.pep	SAQFGPQQIILLLIYAAASAVALFAEPAHIGSLDGTIAWVCFYVCCNLNTLIGYGSFGEAL	240
	orf104ng	SAQFGPQQIILLLIYAAASAVALFLLAEPAHIGSLDGTIAWVCFYVCCNLNTLIGYGSFGEAL	240
35	orf104.pep	KHWEASKVSATVTLLPVFTVINTLLGHVYVMPETFAAFAAEDMNGLGYAGALVVVGAVTAAVG	277
	orf104ng	KHWEASKVSATVTLLPVFTVIFSLLGHVYVMPDTFAAFAAEDMNGLGYAGALVVVGAVTAAVG	300

The complete length ORF104ng nucleotide sequence <SEQ ID 405> is predicted to encode a

40 protein having amino acid sequence <SEQ ID 406>:

1	MENQRPLLGFLALLAAMTWGTLPLXSVRQLKFVDAPTLVWWRFTVAAAVLFVLLALGGR
5	LFVLALGGR SFRLRDRFSWH SFRLRDRFSWH GISANFVLAIAQGLHYISPTT
10	TQVLWQISFPTMIVVGVLVF KDRMTAAQKIGLVLLLGVLMFFNDKGELSGLGRAYKGVLCAAAGSMAWVCNRAVKLL
15	SGLGGTARGVY LCCAGSMEAN VCYAVAQKLL SAQFGPQQIILLYAAASAVAL
20	FLLXAEPAHF GSLDGTIAWV CFYVCCNLNTLIGYGSFGEAL KHWEASKVSATVTLLPVFTVIFSLLGHVYVMPDTFAAFAAEDMNGLGYAGALVVVGAVTAAVG
25	DREFKR*

Further work revealed the complete gonococcal nucleotide sequence <SEQ ID 407>:

50	1	ATGGGAAACCC AABAGGCCGT CCTAGGCTTC CGCTTGGCAC TTTTGGCGGC
	51	GATGAGCTGG GGGACGCTGC CGATTGCGCTG CGGGCAGCTGA TTGAAGCTTG
	101	TCCATGCGCC GACCGCTGGT TGGGTGCGTT TTACCGTGGC GCGCGCGGTA
	151	TTCTTGTGTT TGCCTGCAATT GGGCGGGCGC CTGGCGAACG GCGGGGATT
	201	TCTCTGCGAT TCATTCAGGC TCTGCTGCTG CGGGCGTGACG GGCAATTTCGG
	251	CAAACTTGTG GCTGATTTGC CAAGGGCTGC ATTATATTTC GCGGACCACAG
	301	ACGGCAGTTG TTGGCGCAT TTGGCGGTTC AGCGATGATTTG TTGTCGGCGGT
	351	GTGGCTGTTT AAAGACCGGA tgACTCGGCC GCAGAAACATG GTTGGTGGTT
	401	TGCTGCTggT CGGGTtgcTT ATGTTTttta ACGGCAAAATT CGGGAGATGTT
	451	TGGGGTTGG CGCCGCTATGC GAAGGGCGTG TTGCTGTG TG CGCGCAGCAG
	501	TATGGCTGG TTGTGTTATG CCGCTGGCCA AAAGCTGCTG TCGGGCCCAT
	551	TCGGGCCCCA ACAGATTCCTG CTGTTGATTG ATGGCGCaag tgccgcTG
	601	TTCCGcggtT TTGccgaacC GGCACACATC GGAAGTTtgg acGGTACGtt
	651	GGCGTGGTT TGTTCCTGTT ATTCGTCGTT GAATACCTTA ATCGGGTACG

701 GCTCGTTCGG CGAGGGCGTTG AACACATTGGG AGGCTTCCAA AGTCAGCGCG
 751 GTAACACACT TGCCTCCCCGT GTTTACCGTA ATATTTTCTT TGCTCGGGCA
 801 TTATGTGATG CCTGTGATCT CCTGGCGGCC GSATATGAAC GGTTGGGTT
 851 ATGTCGGCGC ACTGGTGTG TGCGGGGGTG CGGTTACGGC GCGGGTGGGG
 901 GACAGGGCGT CAAACCGCCG CTAG

This corresponds to the amino acid sequence <SEQ ID 408; ORF104ng-1>:

1 MENQRPLLGF ALALLAAMTW GTLPIAVRQV LKVFVDAPLITV WWRFTVAAA
 51 LFVLLALGGR LPKRRDFSWHSFRLLLGVT GISANFVLLIA QGLHYISPTT
 101 TQVLWQISPF TMIVGVLFV KDRMTAAQKLI GLVLLVGLL MFENDKFGE
 151 SGILGAYAKGV LLCAGSMMAW VCYAVAQKLL SAQFGPQQL LLYIYASAAV
 201 FLFEEPAHII GSDLGTLAWV CFVYCCNLTL IGYGSFGEAL KHWEEKSVSA
 251 VTFLLPVFTV FSLLGHYVM PWFIAFPDMN GLGIVGALVV VGGAVTAAVG
 301 DRFFKRR*

ORF104ng-1 and ORF104-1 show 97.5% identity in 277 aa overlap:

15	10 20 30 40 50 60		
	MENQRPLLGF ALALLAAMTW GTLPIAVRQV LKVFVDAPLITV WWRFTVAAA orf104-1.pep LFVLLALGGR LPKRRDFSWHSFRLLLGVT GISANFVLLIA QGLHYISPTT orf104ng-1 MENQRPLLGF ALALLAAMTW GTLPIAVRQV LKVFVDAPLITV WWRFTVAAA <td></td> <td>10 20 30 40 50 60</td>		10 20 30 40 50 60
20	70 80 90 100 110 120		
	LPKRRDFSWHSFRLLLGVT GISANFVLLIA QGLHYISPTT QVLWQISPF TMIVGVLF orf104-1.pep LPKRRDFSWHSFRLLLGVT GISANFVLLIA QGLHYISPTT QVLWQISPF TMIVGVLF orf104ng-1 LPKRRDFSWHSFRLLLGVT GISANFVLLIA QGLHYISPTT QVLWQISPF TMIVGVLF		70 80 90 100 110 120
25	130 140 150 160 170 180		
	KDRMTAAQKIGLVLLLAGLMMFFNDKRGELSGLGAYAKGVLLCAAGSMMAWCVYAVAQKLL orf104-1.pep KDRMTAAQKIGLVLLLAGLMMFFNDKRGELSGLGAYAKGVLLCAAGSMMAWCVYAVAQKLL orf104ng-1 KDRMTAAQKIGLVLLLAGLMMFFNDKRGELSGLGAYAKGVLLCAAGSMMAWCVYAVAQKLL		130 140 150 160 170 180
30	190 200 210 220 230 240		
	SAQFGPQQILLIYAASAAVFLPFAEPAHIGSLDGTIAWVCFVYCCNLTLIGYGSFGEAL orf104-1.pep SAQFGPQQILLIYAASAAVFLPFAEPAHIGSLDGTIAWVCFVYCCNLTLIGYGSFGEAL orf104ng-1 SAQFGPQQILLIYAASAAVFLPFAEPAHIGSLDGTIAWVCFVYCCNLTLIGYGSFGEAL		190 200 210 220 230 240
35	250 260 270		
	KHWEEKSVAVTITLLPVFTVIXXLLGHYVMPETFAAP orf104-1.pep KHWEEKSVAVTITLLPVFTVIXXLLGHYVMPETFAAP orf104ng-1 KHWEEKSVAVTITLLPVFTVIFSLLGHYVMPDFTAFAFDNGLGLYVGALVVVGGAUTAAVG		250 260 270 280 290 300

In addition, ORF104ng-1 shows significant homology with a hypothetical *H.influenzae* protein:

45 gi|1573895 (27269) hypothetical [Haemophilus influenzae] Length = 306
 Score = 237 bits (598), Expect = 8e-62
 Identities = 114/280 (40%), Positives = 168/280 (59%), Gaps = 8/280 (2%)
 Query: 30 QRXXXXXXXXXXXXXXMFTGTLPIAVRQV LKVFVDAPLITV WWRFTVAAA
 Q+P M WG-LPIA+-QWL +A+ T+V+ P+
 Sbjct: 3 QQPLLGFPTFALITAMAWGSLPIALKQVLQSVNMAQTIVWYRFIIAVSLLAIAYKKQLPE 62
 Query: 89 --KRRDFSWHSFRLLLGVT GISANFVLLIA QGLHYISPTT QVLWQISPF TMIVGVLF 146
 K R +W +L+GV G+++NF+L + L+YI P+ Q+ +S F M++ GVL+F
 Sbjct: 63 IMKVQYAW---IMLIGVIGLTSNFLFSSLNLYIEPSVAQIFIHLSFGMLICGVLF 118
 Query: 147 KDRMTAAQKIXXXXXXXXXXMFNDKRGELSGLGAYAKGVLLCAAGSMMAWCVYAVAQKLL 206
 K+++ QKI +FNFND=F +G+ Y+ GV+L G+ W Y +QKL+
 Sbjct: 119 1E9KLHLHQKIGLFLFLLIQGLFNFNDRDFAFAGLNQYSTGVILGVGGALIWWAYGMQKLM 178
 Query: 207 SAQFGPQQILLIYAASAAVFLPFAEPAHIGSLDGTIAWVCFVYCCNLTLIGYGSFGEAL 266
 +F QQILLI+Y A +P A+ + L IA +CF+YCCNLTLIGYGSFGEAL
 Sbjct: 179 1E9KNSQQILLIEMYLGCAATAFMPMDFSQVQEILT-PLALICFYCCNLTLIGYGSYAEAL 237
 Query: 267 KHWEEKSVAVTITLLPVFTVIFSLLGHYVMPDFTAFAFDNGLGLYVGALVVVGGAUTAAVG 306

W+ SKVS V TL+P+FT++FS + HY P FAAP++N
 Sbjct: 238 NRWDWSKVVITLVLPLFTILFISHIAHYFSPADFAAPELN 277

Based on this analysis, including the presence of a putative leader sequence and several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from

5 *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 48

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 409>:

```

10      1 ATGGTAGCTC GPGGGCTCA TAACCAGAG GPCGATGGT CGAATCCGT
      51 .CCCCAACAA TAATTCAA CCCCCTCGGT CAATGCCAGG GG GTTTGT
      101 T TTGCTGT TTTCTGTTTC CTCCTTCCTC CGGCCCTCCGT TTTTGCCTG
      151 ATTTCTTC CGGGCGAAAT ATCGGAACGG CAGACCGGG TCCTGTTGCG
      201 GTTGCATAAT CAGCGCTGG GCTCATACATC TCCCGATTTG TCTCTAAGAA
      251 AGCCACCAT GCGCACCTG CGTTTACCO AATCCGTCAG CAACAAAGAC
      301 CTTGATGCTC TTGTCGAGTG GGCAAAAGCA AGTTACGGTG CAGAAAGTGT
      351 CTGGAAACCG CTGTATCGTA ACGCTGyCC TTGGGCAAC CTGTCGGGG
      401 AATGGGTGGA AGCGGTSsma AAAGACTGGG AGGCAGGCTG CyGGGAGTCT
      451 TCAGACGCCA TTCTTCTGGA TgCGGAGGG TggCtGATA TggCGGAGc
      501 cTTACGACCATC CTCCCGCTCG GTTGGGACTC TGGGGGCTG TTGGACGgT
      551 GGGCGAACAGA GTGTTTCGAG CTGACCGACCG GCGGCCGCA CCCTCTGTC
      601 ACCTCTGAAc GGGCGGTTT mCTCTCTKTCG GGACTGCTCA GCGGGGCGGT
      651 CCATCTAACG GGTCTGACCC AAATCGGACCG CGGATGCGAT TCTCTGATAG
      701 GCAGGGCGAG TCCCGCAAAA GCACTGCGATC CAAACAAACT CGACAAATACT
      751 rCCGGCGGCC GTTGTTCGCG CGGGAAATG CGCTCTGAAG CGCTGTGTCG
      801 CGAAAGCAGC GAAGAGACCG GTTGGATATA AACGCTGcTT CCCTCATCC
      851 CCCGGTATC GCACTGCGA AGCTCGGCT CGCTGACCC GGGTGTACAC
      901 AATGAATCC TGTATGTTT CGATGCCGTC CTGGC...
  
```

This corresponds to the amino acid sequence <SEQ ID 410; ORF105>:

```

30      1 MVARRAHNPV VVGNSNPXPAT XFQTPRFNAE XLVXLPSVCF LFPAAVSFCR
      51 IFLPAIISER QTAVCLRLQI QAWLQLSSAL SSRKPTMLPTV RFTESVSKQD
      101 LDALWEWAKA SYGAECSWT QLYNLQKSLN LSLPEWVERVX KDWEGAXCES
      151 SDGILADAO WPDMGGRLRQH LALGWHCAGL LDGWRNECTD LTDDGGNNPLF
      201 TLERAXXRPX GLLSRAVHLN GLTSDGRWHII FWIGRRSPHK AVDPNKLNDNT
      251 XAGGVSGGEN PSEAVCRESS EEAGLDKTLL PLIRPVSQLH SLRSVSRGVH
      301 NEILYVFDAV LP...
  
```

Further work revealed the complete nucleotide sequence <SEQ ID 411>:

```

40      1 ATGCCGACCG TCGCTTTTAC CGAATCCGTG AGCAAAACAG ACCTTGTAGC
      51 TCTGTCGAG TGCGCAAAAG CAAGTTACCG TCAGAAAGT TGCTGGAAA
      101 CGCTGTATCTA GAAACCGCTA CTTTTGGGCA ACCTCTGCGC GGATATGGTG
      151 GAACCGCTCA AAAAACAGTC GGAGGGCAGG TGCTCGGAGT CTTCAGACGG
      201 CATTCTCTG GATTCGGACG GCTGGCTGTA TATGGGGGGA CGCTTACAGC
      251 ACTTCGGCTC CGGTGGCGCA TGTCGGGGG TGTTGGACCG CTGGCGCAC
      301 GAGTGTTCG ACCTGACCGA CGGGCGGGCG AACCCCTTGT TCACGCTGGA
      351 ACCGGCGCTC TTCCGGCTCT CGCGACTGCT CGGGCGGCC GTCCATCTCA
      401 ACGGCTGAC CGAACCTGGG GGCGCATGGG ATTTCCTGGAT AGGCAGGGC
      451 AGTCGGCACAA AAGCAGTCGA TCCCAACAAAT CTGCACAAATA CTGCCGGGG
      501 CGGTGTPTCC GGGCGGGAAA TGCGCTCTGA AGCCGTTGTCG CGCGAAAGCA
      551 GCGAAGAACG CGGTGTTGGAT AARACGGCTGC TTCCGGCTAT CGGGCGGTAA
      601 TCCCGACTGC ACAGCGCTGG CCTCGCTGAG CGGGGGTGTAC ACATGAAAT
      651 CCTGTATGTA TTCCGAGGCCG TCTCGGGCA AACTCTCTG CTCTGAAAATC
      701 AGGGATGGA AGTGGGGGGT TTTCGAGAAA TGGAATCCTGG CGGTCTGTTG
      751 GATGCCATGT GTGCGGAAA CATGATGCCAC GACGGCGAAC TTGTTACGCT
      801 GGACCGGTT TGCCGTAGG GTCTGATGGA TGCGCGCCAT CGCGCTGTCCG
      851 AGTGGCTGGA CGGGCATACGT TTATAG
  
```

55 This corresponds to the amino acid sequence <SEQ ID 412; ORF105-1>:

-255-

```

1 MPTVRFTESV SKQDLDALFE WAKASYGAES CWKTLYLNGL PLGNLSPWEV
5 51 ERVKKDWEAG CSESSDGIPL NADGWPDMGR RLQHLALGWH CAGLLDGWRN
101 ECFDLTDGGG NPLFTLERA FRPFGLLSRA VRHNGLTESTD GRWHRFWIGRR
151 SPHKAVDPNK LDNTAAGGV5 GGEMPSSEAVC RESSEEEAGLD KTLLPLIRPV
201 SQLHSRLRSVS RGVHNEILYV FDAVLPTF1 PENQDGEVAG FEKMDIGGLL
251 DAMLSGNNMR DAQLVTLDAF CRYGLIDAAB PLSEWLGDGR L*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N. meningitidis* (strain A)

ORF105 shows 89.4% identity over a 226aa overlap with an ORF (ORF105a) from strain A of *N.*

10 *meningitidis*:

orf105.pep	60	70	80	90	100	110
	ISERQTA VCLRLQ TQAVW LQSSAL SRSR KPTM PTM VRF TE SVSK QDLD ALF EWA KAS YGAES					
orf105a						
				10	20	30
orf105.pep	120	130	140	150	160	170
	CWKTLYLN GXP GLNLSPWE VVERVX KWD WEAG CXE SSDG IF IFL NADG WP DMGR LQH LAL GWH					
orf105a	CWKTLYLN GPL GLNLSPWE AERV KK DWE AGC SESS DG IF IFL NADG WP DMGR LQH LAR IWK	40	50	60	70	80
orf105.pep	180	190	200	210	220	230
	CA GLL DG WNR CF DLT DGG GNPL FT LER XA XRP X PG LLS RAV HLN GLT ESD GRH WFWI GRR					
orf105a	EAG L: LH GRW DE CFD LT DGG SNPL FALERA AF PFG LLS RAV HLN GLT ESD GRH WFWI GRR	100	110	120	130	140
orf105.pep	240	250	260	270	280	290
	SP HK AVD PN KLD NT XAG VSG GEM PSE AV CRES SEE AG LDK TLL PL IR PV S QL HS RL RS VS					
orf105a	SP HK AVD PKD LND NT AAG VSG GEL PSE TV CRES SEE AG LDK TLL PL IR PV S QL HS RL RS VS	160	170	180	190	200
orf105.pep	300	310				
	RG VHN EIL YV FD AV LP					
orf105a	RG VHN EIL YV FD AV LP	220	230	240	250	260

40 The complete length ORF105a nucleotide sequence <SEQ ID 413> is:

```

1 ATGCCGACCG TCGCTTTTAC CGAACCTCGT AGCAAACACG ACCTTGATGC
51 CCTATTCCG TGGCCAAGG CAAGTTCAGG TGCGGAAGT TGCTGAAAAA
101 CGCTTATCTTAC GAACGCTCTC CTTTGGGCA CTCTTGCGGT GGAAATGGGG
151 GAGCCGCTCA AAAAGACTCA GGAGGCCAGG TGCTCGAGG CTTCAGACGG
201 CATTTCCTCTG ATATGCCAGC TTCCGGCAGA TAGGGCCAGA CGCTTGAGC
251 ACCTGGCCCG ATATCTGAAA GAAGCGGGAC TGCTTCAGG CTGGGGCAC
301 GAGTTTGGT ACCCTGGGAA CGCCGGCAGC ATTCCTTGT TCCGGCTCGA
351 ACCGGCGCTT TTCCGGCTG CAGGACTGCT CAGCCGGCC GTCCATCTCA
401 ACGTTTGGT CGAACGGGT GGCGCATGGT ATTTCCTGGAT AGGCAGGGC
451 AGTCGGCACA AAGCAGCTGA TCCCAGAAA CTGACATA CTGGCCGGG
501 CGGTGTTTCC AGGGTGTGAA TGGCGCTCTGA AACCGTGTGT CGCGAAAGCA
551 CGCGAAGAACG CGGTGTTGGAT AAAAGCTGCG TTCCGCTCAT CGCCCGGTTA
601 TCCGCGCTGC ACAGCTCTGG CCCCGTCAGG CGGGGTGTGC AACATGAAAT
651 CCTGTATGTA TTCCGATGGC TCTCTGGCGA AACCTTCTG CCTGAAATC
701 AGGATGGCA AGTGGGGGT TTGGAGAAA TTGGACATCGE CGGTCTGTG
751 GCTGCCATGT TGTGGAAA CATGATGCGA GACGGCCACAC TGGTACGCT
801 GGACCGCTT TGGCGTTAGC GTCTGTGA TGCGGCCCAT CGCGTGTGCG
851 AGTGGCTGGA CGCGCATACGT TIAATG

```

This encodes a protein having amino acid sequence <SEQ ID 414>:

60 1 MPTVRFTESV SKHLDLDALFE WAKASYGAES CWKTLYLNGL PLGNLSPWEA
51 ERVKKDWEAG CSESSDGIPL NADGWPDMGR RLQHLALGWHD EAGLHNGWHD

101 ECFDLTDGGS NPLFALERAA FRPFGLLSRA VHNLGLVESD GRWHFWIGRR
 151 SPHKAVDPDK LDNNTAAGGVS SGELPSETVC RESSEEAQGLD KTLLPLIRPV
 201 SQLHSLRVPV RGVRINELLYV FDALVPLTFL PENQDGEGAV FEKMDIGGLL
 251 AAMLSGNMMR DAQLVTLDGF CRYGLIDAAH PLSEWLDGIR L*

5 ORF105a and ORF105-1 show 93.8% identity in 291 aa overlap:

	orf105a.pep	10	20	30	40	50	60	
	MPTVRFTEVKSKHLDLDALEWAKASYGAESCWKTLYLNGLPLGNLSPWAERVKKDWEAG							
10	orf105-1	MPTVRFTEVKSKQDLDALEWAKASYGAESCWKTLYLNGLPLGNLSPWAERVKKDWEAG	10	20	30	40	50	60
	orf105a.pep	70	80	90	100	110	120	
15	orf105-1	CSESSDGFILNADGWPDNGRRLQHLARIKEAGLHHGWRDECFLDTGGSNPLFALERAA						
		CSESSDGFILNADGWPDNGRRLQHLALGWHCAGLDDGWRNECFDLTGCGGNPLTLLERA	70	80	90	100	110	120
	orf105a.pep	130	140	150	160	170	180	
20	orf105-1	FRPFGLLSRAVHLNLVESDGRWHFWIGRSPHKAADVDPDKLNTAAAGGVSSGEPLSETVC						
		FRPFGLLSRAVHLNLTESDGRWHFWIGRSPHKAADVDPDKLNTAAAGGVSSGEMPSAEAVC	130	140	150	160	170	180
	orf105a.pep	190	200	210	220	230	240	
25	orf105-1	RESSEEAQGLDKTLPLIRPVSQLHSLSRVRSPGVHNEILYVFDALVPLTFLPENQDGEVAG						
		RESSEEAQGLDKTLPLIRPVSQLHSLSRVRSPGVHNEILYVFDALVPLTFLPENQDGEVAG	190	200	210	220	230	240
	orf105a.pep	250	260	270	280	290		
30	orf105-1	FEKMDIGGLLAAMLGSNNMHDQAQLTLDLAFCRYGLIDAAPLSEWLGDGIRLK						
		FEKMDIGGLLAAMLGSNNMHDQAQLTLDLAFCRYGLIDAAPLSEWLGDGIRLK	250	260	270	280	290	
35	<u>Homology with a predicted ORF from <i>N.gonorrhoeae</i></u>							

ORF105 shows 87.5% identity over a 312aa overlap with a predicted ORF (ORF105.ng) from *N. gonorrhoeae*:

40	orf105.pep	MVARRAHNPKVVGNSNPXPATXQTPTRPNAEXVLXLPVSCFLFPAAASFVCRIFLPLAISER	60
	orf105ng	: : : : : : : :	
	MVARRAHNPKVVGNSNPAPATKYQTPTRPNAEGVLF-----FLFPAAASFVCRIFLPLAISER		55
45	orf105.pep	QTAVCLRLQIQAQVWLQSSALSSRKPTMPVTRFTESVKQDLDLAEWAKASYGAESCWKT	120
	orf105ng	: : : : : : : :	
	QTAVCLRLQIQAQVWLQSSALCSRKPAAMPVTRFTESVKQDLDLAFERAKASYGAESCWKT		115
50	orf105.pep	LYLNGXPLGNLSPWEVERVXWDWEAGCXESSDGFILNADGWPDMGGRLQHLALGWCHAGL	180
	orf105ng	: : : : : : : :	
	LYLNRPLPLGNLSPWEAERIKDWEAGCSEESNSGIFLNAJDGWPDMGGRLQHLARTWNKAGL		175
55	orf105.pep	LGGWRNECCFDLTDGGNPLFTLERAXXRPKGLLSRAVHLNLTESDGRWHFWIGRSPHK	240
	orf105ng	: : : : : : : :	
	LGGWRNECCFDLTDGGNPLFTLERAAFRPFGLLIRPVSQLHSLSRVRSPGVH		235
60	orf105.pep	AVDPNKLNDNTXAGGVSGGEMPSEAVCRESSEAGLDKTLPLIRPVSQLHSLSRVRSPGVH	300
	orf105ng	: : : : : : : :	
	AVDPGKLDNIAAGGGVSGGEMPSEAVCRESSEAGLDKTLPLIRPVSQLHSLSRVRSPGVH		295
	orf105.pep	NEILYVFDAVLP	312
	orf105ng	NEILYVFDAVLPETFLPENQDGEVAGFERMDIGGLLDAMLSNNMHDQAQLVTLDAFYRYG	355

A complete length ORF105ng nucleotide sequence <SEQ ID 415> was predicted to encode a protein having amino acid sequence <SEQ ID 416>:

1 MVVARAHNNPK VVGSNPAPAT KYQTPRFNAE GVLFFLFPAA SVFCRIFLPA
 51 AISERQAAVC LRLQIAQAVWL QSSALCSRKRP AMPITVRFTES VSKQDLDALF
 101 ERAKASYGAE SWCSTLYLNRL PLGNLSPWE AERIJKDWEA GCSESSNGIF
 151 LNADGWPMG GRQLHILARTW NKAGLLHGWRN NECFDLT/DGG GNPLFTLERA
 201 AFRPFGILLIR AVHLNLGIVES NGRWHFWIGR RSPHKAVDVG KLDNIAGGGV
 251 SGEMPSEAV CRESEEAGL DRTLFLIRP VRSLHSLRPV SRGVHNEILY
 301 VFDAVLPETF LPENQDGVEA GFKEKMDIGGL LDAMLSKNMM HDQLVTLDA
 351 FYRYGLIDAQHPLSEWLGDQI RL*

Further work revealed the complete nucleotide sequence <SEQ ID 417>:

10 1 ATGCCGACCG TCCGTATTAC CGCATCCGTC AGCCAACAAG ACCTTGATGC
 51 CGCTGTGAG CGGCCAAAGT CAAGTTAGCG TGCGGAAAGT TGCTGGAAAAA
 101 CGCTGTATCT GAACCGTCTT CTTTTGGCA ATCTGTGCGC GGATCTGGCT
 151 GAGCGCATCA AAAAGACATG GGAGECAGGC TGCTCOGAGT CTTCAGACGG
 201 CATTTTTCTG AATGCGGAGC GTTGGCGGA TATGGGGGA CGCTTGCAGC
 251 ACCGCGGCCG CACATGGAC AACGGCGGGGC TGCTTACCGG ATGGGGCAC
 301 GAGTGTTCG ACCTGACCA CGGGCGCGGC AACCCCTGTG TCACGCTGGA
 351 ACCGCGCGCT TTCCGTCTG TCGGACTACTA CAGCGCGGGC GTCCATCTCA
 401 ACGGTGTTG CGAATCGAGC CGGAGATGG ATTGTTGGAT AGGCAGGGC
 451 AGTCCGACAA AGAACGTGCG tccggcaac AGTCCGAAATA TTGCGCGGG
 501 CGCTGTTCG GCGCGCGAA TGCGCTCTGA AGCCGTGTGC CGCGAAAGCA
 551 GCGGAAGANG CGGTGTTGGAT AAAAGCGCTGT TTCCGCTCAT CGCCCCAGTA
 601 TCCGGCTGC ACAGCTCTG CCCCTCAGC CGAGGTGTGC ACATGAAAT
 651 CCTGTATCTG TTCTGATGCC CGGGCGGGGA AACCTTCTG CTGAAAAATC
 701 AGGATGGGA GTTGAAGGGT TTGAAAGAA TGACATGG CGCGCTATTG
 751 GATGGCGATG TGTGCAAGGAA CTGATGTCAC GACGGCGAAC TGTTTACGCT
 801 GGACCGTT TACCGTTAGG GTCTGATTTGA TGCGGCCCAT CGCGCTCCG
 851 AGTGGCTGGA CGGCATACGT TTATAG

This corresponds to the amino acid sequence <SEQ ID 418; ORF105ng-1>:

30 1 MPTVRFTESV SKQDLDALFE RAKASYGAE CWKTLIYLNRN PLGNLSPWEWA
 51 ERIKDKWEAG CSESSNGIF NLADGWPMGQ RLQHILARTW NKAGLLHGWRN
 101 ECFDLTDGG NPLFTLERA FRPFGLLSRA VHLNLGVES GRWHFWIGR
 151 SPHKAVDVGK LDNIAAGGVG SGEMPSEAV CRESEEAGL KTLFLIRPV
 201 SRHLHSLRPV RGVHNEILY VFDAVLPETF PENQDGVEAG FEKMDIGGL
 251 DAMLSKNMM HDQLVTLDA YRYGLIDAQHPLSEWLGDQI RL*

35 ORG105ng-1 and ORF105-1 show 93.5% identity in 291 aa overlap:

	10	20	30	40	50	60
orf105-1.pep	MPTVRFTESVSKQDLDALFEWAKASYGAE	CWKTLIYLNRNPLGNLSPWEV	VVKKDWEAG			
orf105ng-1	MPTVRFTESVSKQDLDALFERAKASYGAE	CWKTLIYLNRNPLGNLSPWEAERIJKDWEAG				
	10	20	30	40	50	60
orf105-1.pep	CSESSDGIFLNADGWPMGRLQHILALGHWCAGL	LDGWRNECFDLTDGGGNPLFTLERA				
orf105ng-1	CSESSDGIFLNADGWPMGRLQHILARTWNKAGL	LHGWRNECFDLTDGGGNPLFTLERA				
	70	80	90	100	110	120
orf105-1.pep	FRPFGLLSRAVHLNLGVESNGRNWHFWIGR	RSPHKAVDPGKLDNIAAGGVSGGEMPS	EA			
orf105ng-1	FRPFGLLSRAVHLNLGVESNGRNWHFWIGR	RSPHKAVDPGKLDNIAAGGVSGGEMPS	EA			
	130	140	150	160	170	180
orf105-1.pep	RESSEAGLDKTLPLIRPVSQLHSRLSRSV	GWRVHNELIYVFDAVLPETFPLPENQDG	EVAG			
orf105ng-1	RESSEAGLDKTLPLIRPVSQLHSRLSRSV	GWRVHNELIYVFDAVLPETFPLPENQDG	EVAG			
	190	200	210	220	230	240
orf105-1.pep	FEKMDIGGLDAMLSGNRMHDQLVTLD	AHPLSEWLGDIRLX				
orf105ng-1	FEKMDIGGLDAMLSKNMMHDQLVTLD	AHPLSEWLGDIRLX				
	250	260	270	280	290	

Furthermore, ORF105ng-1 shows homology with a yeast enzyme:

5 sp|P41888|TNR3_SCHPO THIAMIN PYROPHOSPHOKINASE (TPK) (THIAMIN KINASE)
>gi|1076928|pir|S52350 thiamin pyrophosphokinase (EC 2.7.6.2) - fission yeast
(Schizosaccharomyces pombe) >gi|666111|X84417| thiamin pyrophosphokinase
(Schizosaccharomyces pombe) >gi|2330852|gnl|PID|e334056 (298533) thiamin
pyrophosphokinase [Schizosaccharomyces pombe] Length = 569
Score = 105 bits (259), Expect = 4e-22
Identities = 64/192 (33%), Positives = 49/192 (48%), Gaps = 3/192 (1%)

10 Query: 269 NKAGLLHGWRCNECFDLTDGGGNPLFTLERAAFRPPGFLLSRAVHLNLGVESNGRW--HFVII 441
N G+ WRNE + + P+ +ER F FG LS VH + + W+
Sbjct: 96 NTFQIADQWRNELYTVQGSKRKPVLAVERGGCFLSTGVHCTMYIPATKEHPRLRTW 155

15 Query: 442 GRRSPKAVDPGKLDNIAACCCVSGEMPSEAVCRESEEAGLKTFLFLPVRSLHSRL 621
RRSPK P LDN CG++ G+ + +EEA LD + LI P + ++
Sbjct: 156 PRRSPKTQWTNPYLDNSVAGGIAHGSVIGTMKIEFSEEEANLVSSMNLI-PCGTVSYK 214

20 Query: 622 PFDVCP-VHNEILYVDFAVLPETFLPENQDGEFKMDIGGLLDAMLSKNMHHADAQLVT 798
R + E+ YVFD + + +P DGEVAGF + + +L + K+ + LV
Sbjct: 215 MEKRNHIQPELQYVFDLVPDDLVIPRNDGEVAGFSLLPLNQVLHELELKSFKPNCALVL 274

25 Query: 799 LDAYFRYGLIDAAHP 843
LD R+G+I HP
Sbjct: 275 LDFLRHGIITPQMP 289

25 Based on this analysis, including the presence of a putative transmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 49

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID

30 419>:

```

1 ATGAAATGAC CCAAGCAACC CTTCTTCGCT CCCGAAGTGC CCCTTGCCCCG
51 CCAAAACCGC CTGAGCGCTA AACTGATTCT GACACGCCG TTCTCATTTT
103 CCCTATGGC GACATTTGCA TCGATATCTCG CCTTATTTGAT TATCTGTGTT
151 TTGATATTTC GTAACATATAAC CGGAAGAACAC ACAGTGGAGG GACAAATTTT
201 ACCCTCATCG CGCTTAATCA CGGTGTATGC ACCGgATAGC rGKACAAATT
251 CAGCGAAATT CCGTGGAGAGT GGmssAAAAGG TTAAAGCTGG CGACACAGCTA
301 TTTCGGCTTT CGACCTCACG TTTCCGGCGC GGAGGTAGGG TGACAGCAGCA
351 GTTGAACACCG GAGGCAGTTT TAAGAAAACG TTGGCAGAGA CAGGAACCTGG
401 GTCGTCTGAA CCTGATACAGC GGGAAATGAAN CGGCCAGCt TAAGAACATCT
451 GTCGAACTGG TTGAAACACCA GGAACCTCCAT ATTTCCCAAC AGATATGACCG
501 TCAGAAAAGG CGCATTAAGAC TTGGCGGAAGA ATATTTGCGA AAATATCGTT
551 TCCATCCCG .CAATGA

```

This corresponds to the amino acid sequence <SEQ ID 420; ORF107>:

```

45          1 MNRPKQFFR PEVAVARQTS LTGKVILTRP LSFSLNTTFA SISALLIILF
51 LIFONYTRKT TVECQILPAS GVIRVYAPD? XTITPAKFWED GXVKVKGDKL
101 FALSTSRFGA GGSVQQQLKT EAIVLKRILAE QELGRILKH GNETRSLKAT
151 VERLENQELH ISQQIDGQKR RIRLAEMLQ KYRFLSXQ*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

50 ORF107 shows 97.8% identity over a 186aa overlap with an ORF (ORF107a) from strain A of *N.meningitidis*:

-259-

	orf107.pep	MNRPKQPFPRPEVAVARQTSLTGKVILTRPLSFLSLWTTFASISALLIILFLIFGNYTRKT	60
	orf107a	MNRPKQPFPRPEVAVARQTSLTGKVILTRPLSFLSLWTTFASISALLIILFLIFGNYTRKT	
5		10 20 30 40 50 60	
	orf107.pep	70 80 90 100 110 120	
	orf107a	TVEGQILPASGVIRVYAPDTXTITAKFVEDGXKVKGDKLFALSTSRSFGAGGSVQQQLKT	
10		70 80 90 100 110 120	
	orf107.pep	TVEGQILPASGVIRVYAPDTGTITAKFVEDGXKVKGDKLFALSTSRSFGAGGSVQQQLKT	180
	orf107a	130 140 150 160 170 180	
15		EAVLKKTLAEGELGRKLHIGNETRSLSKATVERLENQELHISQQIDGQKRRLAEMLQ	
	orf107a	EAVLKKTLAEGELGRKLHIGNETRSLSKATVERLENQELHISQQIDGQKRRLAEMLQ	
		130 140 150 160 170 180	
20	orf107.pep	KYRFLSLXQK	189
	orf107a	KYRFLSANDAVPKQEMMNVKELLEQKAKLDAYRREEVGLLQEIRTQNLTIXSLPQAAK	
		190 200 210 220 230	

The complete length ORF107a nucleotide sequence <SEQ ID 421> is:

25	1	ATGAATAGAC CCAAGCAACC NTTCTTCCGT CCCGAAGTCG CGCTTGCCCCG	
	51	CCAAACCGC CTGACGGGTA AAGTGTATTAC GACACGACCG TTGTGATTTTT	
	101	CCCTATGGAC GACATTGATCA TCGATATCTG CGTTATTGTAT TATCTCTGTT	
	151	TTGATATTGC GTAATATAC GCGGAAAGACA ACAGTGGAGG GACAAATTTT	
	201	ACCTGGATCG GGCTTAATCA GGGTGTATGC ACCGGATACG GGGCAAAATTA	
	251	CNGGCAAATTCT CNTGGAAAGAT GGAAAAGATA TTAAAGGTGG CGAACAGCTA	
	301	TTTGGGCTTT CGAACCTCAGG TTTCGGGCCA GGAGAGTAGG TGAGCAGAGCA	
	351	GTTGAAAACG GAGGCAGTT TGAGAAAACG GTTGGCAGAGA CAGGAACCTGG	
	401	GTGCTCTGAA GTCTGATACR GGGAAATGAA CGCCGACGCT TAAGGAAACT	
	451	GTGGAACGGT TGGAAAACCA GGAACCTCAT ATTGGCAC AGATGAGACGG	
	501	TCAAGAAAAAGC CGCATAGAC TTGCGGAGA ATATGTCAG AAATATCGTT	
	551	TCTTATCCGC CAATGATGCA GTGCCAAACAC AAGAAATGAT GAATGTCAG	
	601	GCAGACGCTT TAGAGCAGAA AGCCCAACATT GATGCTTACCG GCGGAGAAGA	
	651	AGTCGGGCTC CTTGAGAAA TCCGGAGCA GAATCTGACA TTGGNNAGCC	
	701	TCCCCCAAAGC GGCTATGA	

This encodes a protein having amino acid sequence <SEQ ID 422>:

40	1	MNRPKQPFPRPEVAVARQTS LTGKVILTRPLSFLSLWTTFASISALLIILFLIFGNYTRKF	
	51	LIGNYTRKT TVEGQILPASGVIRVYAPDTXTITAKFVEDGXKVKGDKLFALSTSRSFGAGGSVQQQLKT	
	101	FALSTSRSRFGA GDSVQQQLKT EAVLKKTLAE QPILGRKLHIGNETRSLSKAT	
	151	VERLENQELH ISQQIDGQKR RIRLAEPMIQ KYRFLSANDA VPKQEMMNVK	
	201	AELLEQPKAKL DAYRREEVGL LQEIRTQNLTIXSLPQAA*	

45 Homology with a predicted ORF from *N.gonorrhoeae*

ORF107 shows 95.7% identity over a 188aa overlap with a predicted ORF (ORF107.ng) from *N.gonorrhoeae*:

50	orf107.pep	MNRPKQPFPRPEVAVARQTSLTGKVILTRPLSFLSLWTTFASISALLIILFLIFGNYTRKT	60
	orf107ng	MNRPKQPFPRPEVAVARQTSLTGKVILTRPLSFLSLWTTFASISALLIILFLIFGNYTRKT	60
	orf107.pep	TVEGQILPASGVIRVYAPDTXTITAKFVEDGXKVKGDKLFALSTSRSFGAGGSVQQQLKT	120
	orf107ng	TMEGQILPASGVIRVYAPDTGTITAKFVEDGXKVKGDKLFALSTSRSFGAGGSVQQQLKT	120
	orf107.pep	EAVLKKTLAEQPIGRKLHIGNETRSLSKATVERLENQELHISQQIDGQKRRLAEMLQ	180
	orf107ng	EAVLKKTLAEQPIGRKLHIGNETRSLSKATVERLENQELHISQQIDGQKRRLAEMLQ	180
	orf107.pep	KYRFLSLXQK 188	
	orf107ng	KYRFLSAQ 188	

The complete length ORF107ng nucleotide sequence <SEQ ID 423> is predicted to encode a protein having amino acid sequence <SEQ ID 424>:

5	1 MNRPKQFFFF PEVIAIRQTS LTGKVILTRP LSFSLMTTFA SISALLIILF 51 LIFGNYTRKT TMEGQLPAS GIVRIVYAPDT GTITAKFWED GEKVKRAGDKL 101 FALTSRSRFGA GGVSQQQLKT EAVLTKTLLAE QELGRKLKH ENETRSLKAT 151 VERLENQKLH ISQOIDGGQKR RIRLAAEEMIR KYRFLSAC*
---	---

Based on the presence of a putative transmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

10 Example 50

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 425>:

15	1 ATGCTGAATA CTTTTTTTGC CGTATTGGGC GGCTTGCTGC TGCT TTGCC 51 GTCCGGCAAA TCCGTAATAA CGGGGTGACA GCGCAGAAC GCGGTACAAA 101 GCGCGCCGAA ACCGGTTTC AAAGTCATAT ATATCGACAA TACGGGATT 151 GCGGGTTGG ATTGGGACAA AAGCAGCGAA GGCAAAACCA AGCACGGCAA 201 AAAACAAATC AGTTATCCG TAAAGGCTT GCGCGAACAA ATATGTTATCC 251 GACTGATCGG CAAGCATCCC GGCGACTTGG AAGCCGTCAC CGGCAAATGT 301 ATGGAAACCG ATGATAGGAA CAGTCCGGCA GGTTGGCGA AAAACGGGT 351 GTGCCATACC TTGTTGGCA AACTGTGGG CAATATGCC GAAGACGGCG 401 GCAAACTGAC GGATTACCTA GTTTCGATG CGGCCCTGCAC ACCCTATCAG 451 GCAGGGAAAA CGCGCTGACG ACGGACGCT ATGTGCTTGG 501 AATCGACAGC GAAGGGGGGT TTATTTCCG CGCCCGCCAT TATTGA
----	--

This corresponds to the amino acid sequence <SEQ ID 426; ORF108>:

25	1 MLNTFSFAVLG GCLLXLPGKG SVNTAVOPON AVQSAPKPVF KVVIYIDNTAI 51 AGLDLGSSE GKTNDGKQI SYPIKGLPEQ NVIRLIGKHP GLEAVSGKC 101 METDDKPSA GWAENGVCHT LFAKLVGNTIA EDGGKLTDYL VSHAALQPYQ 151 AGKSGYAAVQ NGRYVLEIDS EGAFYFRRHH Y*
----	---

Further work revealed the following DNA sequence <SEQ ID 427>:

30	1 ATGCTGAAAA CATCTTTTGC CGTATTGGGC GGCTTGCTGC TGCTTGGCC 51 CTGGGGCAAA CCTGGAAATA CGGGGGAAACA GCGGCAAAAC GCGGTACAAA 101 GCGCGCCGAA ACCGGTTTC AAAGTCATAT ATATCGACAA TACGGGATT 151 GCGGGTTGG ATTGGGACAA AAGCAGCGAA GGCAAAACCA AGCACGGCAA 201 AAAACAAATC AGTTATCCG TAAAGGCTT GCGCGAACAA ATATGTTATCC 251 GACTGATCGG CAAGCATCCC GGCGACTTGG AAGCCGTCAC CGGCAAATGT 301 ATGGAAACCG ATGATAGGAA CAGTCCGGCA GGTTGGCGA AAAACGGGT 351 GTGCCATACC TTGTTGGCA AACTGTGGG CAATATGCC GAAGACGGCG 401 GCAAACTGAC GGATTACCTA GTTTCGATG CGGCCCTGCAC ACCCTATCAG 451 GCAGGGAAAA CGCGCTGACG ACGGACGCT ATGTGCTTGG 501 AATCGACAGC GAAGGGGGGT TTATTTCCG CGCCCGCCAT TATTGA
----	---

This corresponds to the amino acid sequence <SEQ ID 428; ORF108-1>:

45	1 MLKTSFAVLG GCLLXLAAAGC SENTAEOPON AVQSAPKPVF KVVIYIDNTAI 51 AGLDLGSSE GKTNDGKQI SYPIKGLPEQ NVIRLIGKHP GLEAVSGKC 101 METDDKPSA GWAENGVCHT LFAKLVGNTIA EDGGKLTDYL VSHAALQPYQ 151 AGKSGYAAVQ NGRYVLEIDS EGAFYFRRHH Y*
----	---

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.gonorrhoeae*

ORF108 shows 88.4% identity over a 181aa overlap with a predicted ORF (ORF108.ng) from *N.gonorrhoeae*:

5	orf108.pep	MLNTFFAVLGGCLLXLPCKSVNTAVQPQNAVQSAPKPVFKVYIDNTAIAGLDLGQSSE	60
	orf108ng	MLKIPFAVLGGCLLLAACGKSENTAEQPQNAQSAPKPVFKVYIDNTAIAGLDLGQSSE	60
10	orf108.pep	GKTNDGKKQISYP1KGLPEQNVRILKGKPGDLEAVSGKCMETDDKDSPEAGWAENGVCHT	120
	orf108ng	GKTNDGKKQISYP1KGLPEQNVRILTGKHPNDLEAVVGKCMETDGKDAPSGWAENGVCHT	120
15	orf108.pep	LFAKLVGNIAEDGGKLTDYLVLVSHALQPYQAGKSGYAAVQNGRYVLEIDSEGAFYFRRRH	181
	orf108ng	LFAKLVGNIAEDGGKLTDYLISHSALQPYQAGKSGYAAVQNGRYVLEIDSEGAFYFRRRH	181

ORF108-1 shows 92.3% identity with ORF108ng over the same 181 aa overlap:

20	orf108-1.pep	MLKTSFAVLGGCLLLAACGKSENTAEQPQNAVQSAPKPVFKVYIDNTAIAGLDLGQSSE	60
	orf108ng-1	MLKIPFAVLGGCLLLAACGKSENTAEQPQNAQSAPKPVFKVYIDNTAIAGLDLGQSSE	60
25	orf108-1.pep	GKTNDGKKQISYP1KGLPEQNVRILKGKPGDLEAVSGKCMETDDKDSPEAGWAENGVCHT	120
	orf108ng-1	GKTNDGKKQISYP1KGLPEQNVRILTGKHPNDLEAVVGKCMETDGKDAPSGWAENGVCHT	120
30	orf108-1.pep	LFAKLVGNIAEDGGKLTDYLVLVSHALQPYQAGKSGYAAVQNGRYVLEIDSEGAFYFRRRH	181
	orf108ng-1	LFAKLVGNIAEDGGKLTDYLISHSALQPYQAGKSGYAAVQNGRYVLEIDSEGAFYFRRRH	181

The complete length ORF108ng nucleotide sequence <SEQ ID 429> is:

30	1	ATGCTGAAA tacctTTTGC CGTGTggc ggCtgctGC TGGTTGCCG	
	51	CTCGCGCAAA TCCGAAAAATA cggcgcaACG GCGCGAAATAA gcgccACAAAA	
35	101	GCGCGCCGAAA ACCGGTTTTC AAACTCAAT ACATCGACAA TAGCGGATT	
	151	GCGCGGGTTT CTTTGGGACA AAAGTGGCAA GGCAGGAAACCA acggcgGCAA	
	201	AAACAAATC AGTTATCcgA TTAAAGGCTT GCGCGAACAA Aacgcgcgtcc	
	251	gGCTGACCCG AAAGCATCCC AACGACTTGG AaggcgctG CGGCGAAATGT	
	301	ATGGAAACCC CGCGGAAGGA CGGGCGCTTG GGCCTGGCGG AAAACCGCGT	
	351	GTGCACTACCC TTGTTGGCA AACTGGTGG CAATATCGCC GAAAGACGGCG	
	401	GCAAACTCG TGATTCCTG ATTTCGATT CGCGCGCTGCA ACCCTATCAG	
	451	GCAGGCCAAA CGGGCTATGC CGCGCGTCG AACGGACGCT ATGTCGTTGA	
	501	AATCGCACCG GggggGGGT TTTtttccg cggccggcat tatttgA	

40 This encodes a protein having amino acid sequence <SEQ ID 430>:

1	MLKIPFAVLG GCCLLXLPCKSVNTA EQPQNAVQSAPKPVFKVYIDNTAIAGLDLGQSSE	
51	AGLALGQSSE GKTNDGKKQI SYP1KGLPEQ NAVRLTGKHP NDLEAVVGK	
101	METDGKDAPS GWAENGVCHT LFAKLVGNIA EDGGKLTDYL ISHSALQPYQ	
151	AGKSGYAAVQ NGRYVLEIDS EGAFYFRRRH Y*	

45 Based on this analysis, including the presence of a predicted prokaryotic membrane lipoprotein lipid attachment site (underlined) and a putative ATP/GTP-binding site motif A (P-loop, double-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

50 **Example 51**

The following DNA sequence was identified in *N.meningitidis* <SEQ ID 431>:

1 ATGGAAGATT TATATATAAT ACTCGCTTGT GGTTGGTTG CGATGATTGC
 51 CGGATTATTC GATGcgatTg cggcggggg TGTTTTGATT ACGGCTGCCG
 101 CACTCTTGTG GCGAGATCT CCTCCCGTGT CGCAATTTGC CACCAACAAG
 151 CTGCAAAGCAG CGCGCTGCTAC GTTTTCAAGT ACGGTTCTT TTGACCGCAA
 5 201 AGGTGTTGATT GATGGAAAAG AAGTCTCCC GATTGCCA GCATCGTTG
 251 TAGGCCGGCT GGCCTGTC TAATCGGTC GCTTGGTTTC CAAAGATATT
 301 CTgCTgCGC TCCTGCCCGT TTGTTGATA TTGTCGCAC TGATTTTGT
 351 GTTTTCCCGC AAAGCTCGACG CGACTAAGGA AGGCAAAAGCC AGAATGCTT
 401 TTTTCTGTG cGGGCTGACG GTCCG ACAG CTTTTGGTTT TTACGACGG
 451 TGTGTCRGGG CGGGCTGTCG GCTCGTTTT TCTGATGTCG TTATTTGTTT
 501 TGCTCGGGTG CAAGCTGGTC AACCCGATGT CTTACACCAA ATGGCGAAC
 551 GTTGGCTGCGA ATCTTGGTC GCTATCGGTA TTCTGCTGCG ACGGTTGCGAT
 601 TATTTTCCCG ATTCGGGCAA CtgTGGGGT CGGTGGCTT GTGGGICGGA
 651 ATTTAGTGC GAGATTGCC GtaAgctTGC GTTCAAGACT GATTRA

15 This corresponds to the amino acid sequence <SEQ ID 432; ORF109>:

1 MEDLYIILAL GLVAMIAFGI DAIAGGGGLI TLPLALLAGI PPVSAIATNK
 51 LQAAATTFSA TVSFARKGLI DWKKGLPIAA ASFVGGVAGA LSVLSLVKDI
 101 LLAVVPVLLI FVALYFVFSP KLDGSKEGKA RMSFFLFLGLT VXTAFGLRR
 151 CVTRGCRVLF SDCLYCFARI QAVERDVLHQ IGERCLQSWF AIGIPAAFRD
 20 201 YFPDCNDGG RCVCFEER EICRTLRFEA D*

Further work revealed the following DNA sequence <SEQ ID 433>:

1 ATGGAAGATT TATATATAAT ACTCGCTTGT GGTTGGTTG CGATGATTGC
 51 CGGATTATTC GATGcgatTg CGGCCGGGG TGTTTTGATT ACGGCTGCCG
 101 CACTCTTGTG GCGAGATCT CCTCCCGTGT CGCAATTTGC CACCAACAAG
 151 CTGCAAAGCAG CGCGCTGCTAC GTTTTCAAGT ACGGTTCTT TTGACCGCAA
 201 AGGTGTTGATT GATGGAAAAG AAGTCTCCC GATTGCCA GCATCGTTG
 251 TAGGCCGGCT GGCCTGTC TAATCGGTC GCTTGGTTTC CAAAGATATT
 301 CTgCTgCGC TCCTGCCCGT TTGTTGATA TTGTCGCAC TGATTTTGT
 351 GTTTTCCCGC AAAGCTCGACG CGACTAAGGA AGGCAAAAGCC AGAATGCTT
 401 TTTTCTGTG cGGGCTGACG GTCCG ACAG CTTTTGGTTT TTACGACGG
 451 TGTTGGCTG CGGGCTGACG GTCCGACCC TTGGGGTTT TTACGACGGT
 501 GCTGGCTGT AAAGCTGGTGA ACGGGATGTC TTACACCAAAT TTGGCGAACG
 551 TTGGCTGCA TCTGGTTCC TCTACGGTAT TCTGCTGCA CGGTGGATT
 601 ATTTTCCCGA TTGGGGCACG ATGGGGCGTC GGTCGGTTG TGCGTGGCGA
 651 TTAGGTGGC AGATTTGGCC TCCCTTCCG TTGGAGCTG ATTAGACGCC
 701 TGCTGATGT CATCAGCATT TGATGCGCTG TGAAATGTT GATAGACGAG
 751 AGAAATCCCG TGATCAGAT GATTGTTCG ATGTTTAA

This corresponds to the amino acid sequence <SEQ ID 434; ORF109-1>:

1 MEDLYIILAL GLVAMIAFGI DAIAGGGGLI TLPLALLAGI PPVSAIATNK
 51 LQAAATTFSA TVSFARKGLI DWKKGLPIAA ASFVGGVAGA LSVLSLVKDI
 101 LLAVVPVLLI FVALYFVFSP KLDGSKEGKA RMSFFLFLGLT VAPLGYD
 151 VFPGVGSFV LIAFTVLLGC KLNNAMSYT LANVACNLGS LSVLFLHGSI
 201 IFPIATMAV GAEFGANLGA RFPAVRGFSKL IKPLLIVIST SMAVKLLEIDE
 251 RNPLYQMIVS MF*

45 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF109 shows 95.9% identity over a 147aa overlap with an ORF (ORF109a) from strain A of *N. meningitidis*:

50	orf109.pep	MEDLYIILALGLVAMIAFGIDAIAGGGGLTLPLALLAGI PPPVSAIATNKLQAAATFSA	10	20	30	40	50	60
	orf109a	MEDLYIILALGLVAMIAFGIDAIAGGGGLTLPLALLAGI PPPVSAIATNKLQAAATFSA	10	20	30	40	50	60
55	orf109.pep	TVSFARKGLIDWKKGTLPIAAASFVGGVAGALSVLSVSKDILLAVVPVLLIFVALYFVFSP	70	80	90	100	110	120
	orf109a	TVSFARKGLIDWKKGTLPIAAASFAGGVVAGALSVLSVSKDILLAVVPVLLIFVALYFVFSP	70	80	90	100	110	120

	130	140	150	160	170	180		
	KLDGSKEGKARMSSFLFGLTVXATFGELRRCVRTGCRVLFSCLYCFARLQAVERDVLHQ							
5	orf109a	KLDGSKEGKARMSSFLFGLTVAPLLGFYDGVPGPGVGSFFLIAFIVLLGCKLNNAMSYTK	130	140	150	160	170	180

The complete length ORF109a nucleotide sequence <SEQ ID 435> is:

```

1  ATGGAAGATT TATACATAAT ACTCGCTTGC GGTGGTTTG CGATGATTGC
51  CGGATTATTC GATCGGATTC CGGGTGGGGG TGTTTTGATT ACGGTGCCTG
101 CACTCTTGTG GGCAGGTATT CCTCCGCTGT CGGCAATTGC CACCARCAA
151 CTGCACGCG CGCGCTGCTAC TTTTGCGCT GCGGTTTCTT TTGACCGAA
201 AGCTTGATT GATTGGRANGA ARGGCTTCCC GATTGCGCGA GCATCGTTTG
251 CAGGGCGCGT GGTGCGTCA TTATCGCTCA GCTTGGTTTC CAAGAGTATT
301 CTGCTGGGG TCGTGCCTGTT TTGTTGATA TTGTCGCGC TGTTTTTGT
351 GTTCTGGCC AARGCTGACG GCAGTAAGGG AGGCAAGGCC AGAATGCTT
401 TTTTCTGTT CGGTCTGACG GTTGCACAC TTTGGTTT TTACACCGT
451 GTGTCGAGC CGGGTGTGG CTGCTGTTCTT CTGATGCGCT TTATGTTTT
501 GCTGGCGTC AAGCTGTTGA ACGGATGATC TTACACAAA TTGGCGAACG
551 TTGCTCGCACT TTGTTGTTTG CTATCGTAC TCTCGTGCAGA CGGTTGATT
601 ATTTCGGCACT TTGCGGCAAC GATGGCGTC GGTGCGTTG TCCTGGCGAA
651 TTTAGCTGCG AGATTTGCGC TCCGGCTTGG TTCCGAACTG ATTAAAGCCGC
701 TGTGATTTGT CATCAGCAT TCGATGCGT TGAARATGTT GATAGACGAG
751 AGAAATCCGC TGATCAGAT GATTGTTTCG ATGTTTAA

```

This encodes a protein having amino acid sequence <SEQ ID 436>:

```

25  1 MEDLYIILAL GLVAMIAGFL TIPALLAGI PPVSIAITNK
51  51 LQAAATFSA TVSFARKGLI DWKGLPLIAA ASFAGGVVGVA LSVLSVSKD1
101 101 LLAVVVFVLLI FVALFYVFPSP KLDGSKEGKA RMSFFLFLGTVAPLFLGFDG
151 151 VFGPGVGSFPI LIAFIVLGLC KLLNAMSYTK LANVACNLGS LSVFLLEHGS1
201 201 IFFIAATMCAV GAFVGANLGA RFAVRPGSKL IKPLLIVISI SMAVKLLIDE
251 251 RNFLYQMVTS MF*

```

ORF109a and ORF109-1 show 99.2% identity in 262 aa overlap:

	10	20	30	40	50	60		
	MEDLYIILALGLVAMIAGFLDIAAGGGGLITLPALLAGIIPPVSAITTNKLQAAAATFSA							
35	orf109-1	MEDLYIILALGLVAMIAGFLDIAAGGGGLITLPALLAGIIPPVSAITTNKLQAAAATFSA	10	20	30	40	50	60
	70	80	90	100	110	120		
40	orf109a.pep	TVSFARKGLIDWKKGGLPIAAASFAGGVVGALSVSLSKDLILLAVVPVLLIFVALFYVFPSP						
	orf109-1	TVSFARKGLIDWKKGGLPIAAASFAGGVVGALSVSLSKDLILLAVVPVLLIFVALFYVFPSP	70	80	90	100	110	120
	130	140	150	160	170	180		
45	orf109a.pep	KLDGSKEGKARMSSFLFGLTVAPLLGFYDGVPGPGVGSFFLIAFIVLLGCKLNNAMSYTK						
	orf109-1	KLDGSKEGKARMSSFLFGLTVAPLLGFYDGVPGPGVGSFFLIAFIVLLGCKLNNAMSYTK	130	140	150	160	170	180
	190	200	210	220	230	240		
50	orf109a.pep	LANVACNLGSLSVFLLNGSIIFFIAATMAVGRFVGANLGRARFAVRPGSKLIKPLLIVISI						
	orf109-1	LANVACNLGSLSVFLLNGSIIFFIAATMAVGRFVGANLGRARFAVRPGSKLIKPLLIVISI	190	200	210	220	230	240
	250	260						
55	orf109a.pep	SMAVKLLIDERNFLYQMVSMFK						
	orf109-1	SMAVKLLIDERNFLYQMVSMFK	250	260				

Homology with a predicted ORF from *N.gonorrhoeae*

ORF109 shows 98.3% identity over a 231aa overlap with a predicted ORF (ORF109.ng) from *N. gonorrhoeae*:

5	orf109.pep	MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNKLQAAAATFSA	60
	orf109ng	MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNKLQAAAATFSA	60
10	orf109.pep	TVSFARKGLIDWKGLPIAAASFAGGVGVALSVSKDILLAVVPVLLIEVALYFVFS	120
	orf109ng	TVSFARKGLIDWKGLPIAAASFAGGVGVALSVSKDILLAVVPVLLIEVALYFVFS	120
15	orf109.pep	KLDGSKEKGKARMSFFLFLGLTVATAFGFLRRCVRTGCRILVFSCLYCFCARLQAVERDVLHQ	180
	orf109ng	KLDGSKEKGKARMSFFLFLGLTVATAFGFLRRCVRTGCRILVFSCLYCFCARLQAVERDVLHQ	180
	orf109.pep	IGERCLQSWSFAIGIPAARFDYFPDCGNDDGRCVRCEFRCEICRCLRLFEAD	231
	orf109ng	IGERCLQSWSFAIGIPAARFDYFPDCGNDDGRCVRCEFRCEICRCLRLFEAD	231

An ORF109ng nucleotide sequence <SEQ ID 437> was predicted to encode a protein having amino acid sequence <SEQ ID 438>:

1	MEDLYIILALGLVAMIGEL DAIAGGGGLI TLPALLLAGI PPVSAIATN
51	LQAAAATFSA TVSFARKGLID WKGLPIAA ASFAGGVGVA LSVSLVSKDI
101	LLAVVPVLLI EVALYFVFS P KLDGSKEKGK RMSFFLFLGT VAPLIGFLRR
151	CVRTGCRILVF SDCLYCFCARL QAVERDVLHQ IGERCLQSWSF AIGIPAARFD
201	YFPDCGNDDG RCVRCEFR EICRCLRLFEA D*

Further work revealed the following gonococcal DNA sequence <SEQ ID 439>:

1	ATGGAGAAGT TATACATATA ACTCGCTT TGTTTGGTTG CGATGATCGC
51	CGGATTATTC GATGCCGATTC CGGGGGGGGGG TGTTTGGATG AGCGTCGCCTG
101	CACTTGTGTT GGCAAGCTAT CCTCCCGTGT CGGCAATTCG CACCAAACAG
151	CTGCAAGCAG CGCTGCTTA GTTTTGGCTG ACGGTTTCTC TTGCA CGCAA
201	AGGGTTGATT GATTTGGAAA AAAGTCCTCC GATTGCGGCA GCATGTTTG
251	CAGGGGGCTG GGTGCGTGA TTATGGCTCA GCTTGGTTTC CAAGAGATT
301	TTGCTGGCCG TTGTCGGCGT TTGTTGATTA TTTGCTGGCCG TGTATTTTGT
351	GACTTCGCCG AAGCTCGAGC GCACTAAGGG AGGCAAGGG AGATACTCTT
401	TTTTCTATT CGGGCTGAGC GTGACCGG TTTGGTTT TTACGACGGT
451	GTGTTGGCGG CGGGTCTGG CTGTTTTT CTGATTTGCTT TTATTTTTT
501	GCTTCGGCTCG AACGCTGTTGA ACGGCGATGTC TTACACCAA TTGGCGAACG
551	TTGCTGGCCG TTGTCGGCTA CTATGGTAT CCTCTGCTGA CGGTTGATT
601	ATTTCCTCCC CATGAGCTGGT GGTGCGGTO GGTGCGTTG TCCTGGCGAA
651	TTTGGTGGCG AGATTTGGCG TCCGGCTTCG TTGCAAGCTG ATTAAGCGG
701	TGCTGATTTG CATCACGAT CGTATGGCTG TGAATTGTTG GATAGACGGAG
751	AGAAAATCCG TGTATCAGAT GATTGTTTGC ATGTTTTAA

This corresponds to the amino acid sequence <SEQ ID 440; ORF109ng-1>:

45	1 MEDLYIILALGLVAMIGFI DAIAGGGGLI TLPALLLAGI PPVSAIATN
	51 LQAAAATFSA TVSFARKGLID WKGLPIAA ASFAGGVGVA LSVSLVSKDI
101	LLAVVPVLLI EVALYFVFS P KLDGSKEKGK RMSFFLFLGT VAPLIGFLRG
151	VFGPGVGSSF LIRETLVLLGC KLNAMSYTK LANVACNLS LSVFLLHSI
201	IIFTIVATMVA GAFVGANLGA RFAVRFGSKL IKPLLIVISI SMAVKLLIDE
251	RNFLYQHIVS MF*

50 ORF109ng-1 and ORF109-1 show 98.9% identity in 262 aa overlap:

	10	20	30	40	50	60
	orf109ng-1.pep MEDLYIILALGLVAMIGFI DAIAGGGGLI TLPALLLAGI PPVSAIATN					
	101 LQAAAATFSA TVSFARKGLID WKGLPIAA ASFAGGVGVA LSVSLVSKDI					
55	101 LLAVVPVLLI EVALYFVFS P KLDGSKEKGK RMSFFLFLGT VAPLIGFLRG					
	151 VFGPGVGSSF LIRETLVLLGC KLNAMSYTK LANVACNLS LSVFLLHSI					
	201 IIFTIVATMVA GAFVGANLGA RFAVRFGSKL IKPLLIVISI SMAVKLLIDE					
	251 RNFLYQHIVS MF*					
	70	80	90	100	110	120
	orf109ng-1.pep TVSFARKGLID WKGLPIAA ASFAGGVGVA LSVSLVSKDI					
	101 LLAVVPVLLI EVALYFVFS P KLDGSKEKGK RMSFFLFLGT VAPLIGFLRG					
	151 VFGPGVGSSF LIRETLVLLGC KLNAMSYTK LANVACNLS LSVFLLHSI					
	201 IIFTIVATMVA GAFVGANLGA RFAVRFGSKL IKPLLIVISI SMAVKLLIDE					
	251 RNFLYQHIVS MF*					

	orf109-1	: : : : : : TVSFARKGLIDWKKGKGLPIAAASFVGGVAGALSVSILVKSDILLAVFVLLFVALYFVFSP 70 80 90 100 110 120
5	orf109ng-1.pep	130 140 150 160 170 180 KLDSGSKEGKARMSSFLFGILTVAFLLGFYDGVFGFVGGSFFLIAFIVVLGCKLNNAMSYTK
	orf109-1	130 140 150 160 170 180 KLDSGSKEGKARMSSFLFGILTVAFLLGFYDGVFGFVGGSFFLIAFIVVLGCKLNNAMSYTK
10	orf109ng-1.pep	190 200 210 220 230 240 LANVACNLGSLSVFLLHGSIIFPIVATMAVGAFVGANLGRFAVRFGSKLIKPLLVISI
	orf109-1	190 200 210 220 230 240 LANVACNLGSLSVFLLHGSIIFPIVATMAVGAFVGANLGRFAVRFGSKLIKPLLVISI
15	orf109ng-1.pep	250 260 SMAVKLLIDERNPLQYQMVSMFX
	orf109-1	250 260 SMAVKLLIDERNPLQYQMVSMFX

In addition, ORF109ng-1 shows homology to a hypothetical *Pseudomonas* protein:

```
sp|P29942|YCB9_PSEDE HYPOTHETICAL 27.4 KD PROTEIN IN COBO 3'REGION (ORF9)
>gi|94984|pir|I|38164 hypothetical protein 9 - Pseudomonas sp >gi|551929
(M62866) ORF9 [Pseudomonas denitrificans] Length = 261
Score = 175 bits (43%), Expect = 3e-43
Identities = 83/214 (38%), Positives = 131/214 (60%), Gaps = 1/214 (0%)
Query: 41  PPVSAITAKLQXXXXXXXXXXXXXRRKGLIIDWKKGPLIXXXXXXXXXXXXXXXKDI 100
      PP+ + TNKLQ R+G ++ K+ LP+ D+
Sbjct: 43  PFLQTLGTNKLQGLFGSGSATLSYARRGHVNLKEQLPMAIMSAAGAVLGALLATIVPGDV 102
Query: 101 LLAVVPLVLLFVALYFVFSKPLDSKEGKARMSSFLFGILTVAFLLGFYDGVFGFVGGSFF 160
      L+A+F LLI +ALYF P + G + +R+ F+ D+ PL+F+FYDGVFGFPG GSFF
Sbjct: 103 LKALKFPLLIALYFGLPKPNM-GDVKHQSRVTFTLTLVPLIGFYDGVFGFPGTGSFF 161
Query: 161 LIAFIVLGLCKLNNAMSYTKLANVACNLGSLSVFLLHGSIIFPIVATMAVGAFVGANLGA 220
      ++ F+ L G + L A ++TK N N+G+ VEL G++++ + M +G F+GA +G+
Sbjct: 162 MLGVFTLAGFVGVLKATAHTKFLNFGSNVCAFGVLFVFFGAVLWKVGLLMGLQFLGRQWGS 221
Query: 221 RFAVRFGSKLIKPLLVISISMAVKLLIDERNPL 254
      R+A+ G+R+ IKPL++SI++A+LL D +PL
Sbjct: 222 RYAMAKGAKIIKPLLVISIALAIRLLADPTHPL 255
```

Based on this analysis, including the presence of a putative leader sequence (double-underlined)
 45 and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 52

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 441>:

50	1 ..CTGCTAGGGT ATTGCATCGG TTATCGGTAC GGCTGTTGCA GCAAAACAGC
	51 CCGCAGACGG ATTATTTGGT CAATATTCCGA TCTTTTTGGG CGAG ATT TT
55	101 CGGACTTTCTG GGAACTTAAG ACGTCTATGC TTCCGCTATGG TTTGTCGTTA
	151 TCTATGATGTT TTGTTGGTTT TCTAACAGTT TGTCGCTGAT TGCCAAATGTG
	201 CGGCCCTTGT GGGCGGAAT GAAACTTTTT CGGGAAAAGG TAAAGAAAAA
	251 ATCTCTGGG CGCATGGCCC ATTCTTCGGT GTTGGAACTGA AAAATTCGGC
	301 CGGAGGTTGC CAAACGTTAT CTGGAAATAC AAGGTTTTCR GGGAAAAACC
	351 ATTAACCGTG AAAGACGGTC GGTTCGATT GCGCCTAAAA AAGGCACAAAT
	401 AACAAAAATGG GGCTATATCT TTGCCCCATG TGCTTTGATT GTCATTTGCG
	451 TGGCGGGGTG ATGAGACAGT AACCTCGCTG TGAAACTGGG TATGCTGACC
60	501 GGTGCGGATTG TTCCCGACAA TCAGCGGGTT TATGCCAAGG ATTTCAAGC

551 CGCAAAGTAT .TTGGGTGCG gTCCAATCTC TCATTTAGGG GCAACGTCAA
 601 TATTCGG.A GGGGCAAGgt GCGGATGTGG TTTTCCTGA

This corresponds to the amino acid sequence <SEQ ID 442; ORF110>:

5 1 ..LLGIASVIGT LLQQNQPQTD YLVKFGSEWA XIFGFLGLYD VYASAWFVV
 51 MMFLVUSTL CLIRNVPFW REMKSFRKEV KEKSIAAMRH SSLLDVKIAP
 101 EVAKRYLEVQ GFQGKTINRE DGSVLIAAKK GTMNKKGWYIF AHVALIVICL
 151 GGILDNSNLL KLGMLTGRIF RTIIRRMPRI XKPESKFGCV QSLI*GQRQY
 201 FXRGRVRMWF S*

Computer analysis of this amino acid sequence gave the following results:

10 [Homology with ORF88a from *N. meningitidis* \(strain A\)](#)

ORF110 shows 91.5% identity over a 188aa overlap with ORF88a from strain A of *N. meningitidis*:

	10	20	30	40	50	60	
	orf88a.pep	MSKSRSRSPPLLSRPWFAFFSSMRFAVALLSLLGIASVIGT	VILQQNQPQTDYLVKFGSEWA				
15	orf110						
		LLGIASVIGT	LLQQNQPQTDYLVKFGSEWA				
		10	20	30	40	50	
	orf88a.pep	70	80	90	100	110	120
20	orf110	QIFGFLGLYDVYASAWFVIMMFELVUSTLCLIRNVPFWREMKSFRKEVKEKSIAAMRH	: :	: :	: :	: :	: :
		40	50	60	70	80	90
	orf88a.pep	130	140	150	160	170	180
25	orf110	SSLLDVKIAPEVAKRYLEVQGFQGKTINREDGSVLIAAKKGTMNKKGWYIFAHVALIVICL	: :	: :	: :	: :	: :
		100	110	120	130	140	150
	orf88a.pep	190	200	210	220	230	240
30	orf110	GGLIDSNLLKLGMLTGRIVPDNQAVYAKDFKPESELIGASNLSFRGNVNISEGQSADVVF	: :	: :	: :	: :	: :
		160	170	180	190	200	210
	orf88a.pep	250	260	270	280	290	300
35	orf110	LNAADNGILVQDLPFEEVKLKKFHIDFYNTGMPRDFAFS DIEVTDKATGEKLERTIRVNHPLT	: :	: :	: :	: :	: :
		SX					

40 However, ORF88 and ORF110 do not align, because they represent two different fragments of the same protein.

[Homology with a predicted ORF from *N. gonorrhoeae*](#)

ORF110 shows 88.6% identity over a 211aa overlap with a predicted ORF (ORF110.ng) from *N. gonorrhoeae*:

45	orf110.pep	LLGIASVIGTLLQQNQPQTDYLVKFGSEWA	30
	orf110ng	MSKSRSPTLLSRPWFAFFSSMRFAVALLSLLGIASVIGT	60
50	orf110.pep	VILQQNQPQTDYLVKFGSEWA XIFGFLGLYDVYASAWFVIMMFELVUSTLCLIRNVPFWREMKSFRKEVKEKSIAAMRH	90
	orf110ng	: : RIFDFGLGLYDVYASAWFVIMMFELVUSTLCLIRNVPFWREMKSFRKEVKEKSIAAMRH	120
	orf110.pep	SSLLDVKIAPEVAKRYLEVQGFQGKTINREDGSVLIAAKKGTMNKKGWYIFAHVALIVICL	150
55	orf110ng	: : SLLDVKIAPEVAKRYLEVQGFQGKTISREDGSVLIAAKKGTMNKKGWYIXAHVALIVICL	180

-267-

5	orf110.pep	GGLIDSNLLKLGMGLTGRIFRTIIRRMPRIXKPESXFGCVQSLIXQQRQYFXRGRVRMWF	210
		: : : : : : : : : : : : : : : : : :	
	orf110ng	GRLINXNLLKLGMILAGSIFRNRRVMPRISKPESTWGGVQSLIKGQRQYFQRGKVRMWF	240
5	orf110.pep	S 211	
	orf110ng	S 241	

The complete length ORF110ng nucleotide sequence <SEQ ID 443> is predicted to encode a protein having amino acid sequence <SEQ ID 444>:

10	1 MSKSRISPTL LSRPWFAFFS SMRFAVALLS LLGIASVIQT VLQQNQPOTD	
	51 YLVKGFPWT RIFDFLGLYD VYASAWFWV MFLVVTSL CLRNPVPEFW	
	101 REMKSFSREKV KEKSLAMMRH SSSLLEVKIAF EVAKRYLEVR GFGQKTVSRE	
	151 DGSVLIAKKI GTMMNKWGYIX AHALIVICIL GRILINXLLL KLGMILASIF	
	201 RNNRRVMPRI SKPESIWWGV QSLIKGQRQY FQRGKVRMWF S*	

15 Based on the putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 53

The following DNA sequence was identified in *N.meningitidis* <SEQ ID 445>:

20	1 ATCGCGCTG AAACAGCGCT GCCGAACTTT ATCCGGCTCT TGATATTTCG	
	51 CCTGGGTTTC ATCTCTCTGA ACCGCTGTTC GGACACAAACCC GCGCAAAACCG	
	101 TTACCCCTCA AGGGCAAAAGC ATGGGCACAGA CCTTACCGCTT CAATAACCTT	
	151 TCAAATATC CGGACAACAA CCCCTCACCT GCGAAACAT AAAACGAT	
	201 CGATGACGCC CTAAAGAACG TCAACCGGCA GATGTCACCC TATCAGCCUG	
	251 ACTCCGAAT CAGCGCGTTT AACCACACAA CAGCGCGCA GCCCCCTCCGC	
	301 ATTCAACGGC ACTTCGACAA CGTTACTGCG GAAGCCGTCG GCTCTAACCG	
	351 CCTGACACAC GCGCGCGCTG AGCTAAACGGT CGGGCCCTTG GTCAACCTT	
	401 GGGGATTGCG CCCCGAACAAA TCCTGTACCC GTGACCGCTG GCGGGAAACAA	
	451 ATCAAACAGG CGGCATCTTA TACGGCCATA GACAAATACT TTTGAAACA	
	501 AGGCAAAAGAT TACGTGCTCT TGAGCAAAAC CAACCCCCAAAG GCTTATTTGG	
	551 ATTATCTTC GATTCGCAAA GGCTTCGGG TTGATAAAGT TGGGGCGAA	
	601 CTGGAACAAATC AGCGCATCAAA AATTATTCG TGCGAAATCGG CGCGGGAGTT	
	651 GCACGGCCAAA GCGCAAAACCG CGCGGGGGGG ACCGGGGCGG ATCGGTATCG	
	701 AGCGACCCAA TATGTCGCAAA GGCGGCAATA CGCAGATTAT CGTCCCCTGTG	
	751 AACACCGTTT CGCTTGCGCA TTCCGGGAT TACCGTATT TCCACGCTGA	
	801 TAAAAACGGC AAACGGCTT CCCATATCAT CAACCCGAAAC AAAAAACGAC	
	851 CCATCAGCCA CAACTCGGCC TCCATCAGCG TGTCGAGAA CAGTCGGATG	
	901 ACGGCGGAGC GCTTGTGTCAC AGGTTATTCG GTATTTGGGG AAACCGGAGC	
	951 CTAAAGCTG CGAAGACGGG AAAAATCGC TGTTTCTCTG ATTCAGGG	
40	1001 ATAAGGCGG CTACCGOACC GCCATGTCTT CGGAAATTGA AAAACTGCTC	
	1051 CGCTAA	

This corresponds to the amino acid sequence <SEQ ID 446, ORF111>:

45	1 MPSETRLPNF IRVLIFALGF IFLNACSEQT AQTVTLQGET MGTTTYTVKYL	
	51 SNNRKDLSPF AEIKORRIDDA LKEVNRQSTM YQPSESSRF NQHTAGKPL	
	101 ISSDFAHVTA EAVRLNLTH GALDVTVGPI VNLMNGFPDK SVTREPSPEQ	
	151 IKQAASYTGI DKIILKQGKD YASLSKTHPK AYLDLSSIAK GFGVVKVAGE	
	201 LEKYIGIQNYI VEIGGELHKG GKNARGEPRW IGIEQPNIVQ GGNTQIIVPL	
	251 NNRSLATSGD YRIFHVDKNG KRLSHIINPN NKRFISHNL A SISVADSM	
	301 TADGLSTGLF VLGETEALKL AEEKLAFLV IVRDKGGYRT AMSSEFEKL	
50	351 R*	

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF111 shows 96.9% identity over a 351aa overlap with an ORF (ORF111a) from strain A of *N.meningitidis*:

		10	20	30	40	50	60
5	orf111a.pep	MPSETRLPNFIRTLIFALSFIFI	LNACSEQTAQTVTLQGETMGT	TYTVKYL	SNNRDXLPS		
	orf111	MPSETRLPNFIRVLIFALGFI	LNACSEQTAQTVTLQGETMGT	TYTVKYL	SNNRDXLPS		
10	orf111a.pep	10	20	30	40	50	60
	orf111	AEIQKRIDDALEVNRMQSTYQ	PDSEISRFNQHTAGKPLRISSDFAH	HVTAEAVRLNR	LTH		
15	orf111a.pep	AEIQKRIDDALEVNRMQSTYQ	PDSEISRFNQHTAGKPLRISSDFAH	HVTAEAVRLNR	LTH		
	orf111	70	80	90	100	110	120
20	orf111a.pep	130	140	150	160	170	180
	orf111	GALDTVTVGPLVNLWGF	GDPKSVTREPSPBPEQIKQAA	SYTGIDKII	LKQGKD	YASLSKTHPK	
25	orf111a.pep	190	200	210	220	230	240
	orf111	AYLDLSSIAKGF	GFDVKVAGELEKYG	IYLV	EIGGELHGKX	KNARGEPRIGIEQPNI	VQ
30	orf111a.pep	GGNTQII	VPLNNRSXATSGDYR	IFHVDKSGKRLSHI	INPNNKRPISHN	LASIVSVXAD	SAM
	orf111	190	200	210	220	230	240
35	orf111a.pep	250	260	270	280	290	300
	orf111	TADGLSTGLFVLGE	TEALKLAEREKLAVFL	IVRDKG	GYRTAMSSEFE	KLLRX	
40	orf111a.pep	310	320	330	340	350	
	orf111	TADGLSTGLFVLGE	TEALKLAEREKLAVFL	IVRDKG	GYRTAMSSEFE	KLLRX	
45	orf111a.pep	310	320	330	340	350	
	orf111	1	ATGCCGTCG	AAACACGCTC	GGCGAACCTT	ATCCGCACCT	TGATATTG
50	orf111a.pep	51	CCTGGATTTT	ATCTCTCTGA	AGCGCTGTC	GGAAACAAAC	GGCCAAACCG
	orf111	101	TTACCTCTGA	AGGTGAAAGC	ATGSGCAGA	CCTATAACCT	CAAATACCTT
55	orf111a.pep	151	TCAATAATC	GGGACNACT	CCCTCACT	GGCGAAATAC	AAAANGCAT
	orf111	201	CGATGACCGG	CTTAAAGAAG	TCAACCGGG	GATCTCCACC	TATCAGCCCG
60	orf111a.pep	251	ACTCGAAT	CAGCGCGTTC	AACCAACRA	CAGCGCGCMA	GGCCCTCCGG
	orf111	301	ATTCAACGCG	ACTTGCGACA	CGTTACTCGC	GAAGCGCTG	ACCTGAACCG
65	orf111a.pep	351	CCGTGACACAC	GGCGCGCTG	ACGTAACCGT	GGGCCCCCTTG	GTCAACCTTT
	orf111	401	GGGGATTTGCG	CCCCGACAAA	TCGGTATCCO	GTGACCGTC	GGCGAACAA
70	orf111a.pep	451	ATCAAACAAAG	CAGCATCTTA	TACGGCCATA	GACAAATATCA	TTTGAAACAA
	orf111	501	AGGCAAAAGAT	TACCGTTCT	TGAGCAAAAC	CCACCCCAAG	GCCTATTG
75	orf111a.pep	551	ATTATATCTTC	GATGGCCAA	GGCTTCGGCG	TTGATNANT	TGGGGCGCAA
	orf111	601	CTGGAAAAAT	ACGGCATCTA	AAATATTCAT	GTGCAAATCG	GGCGNGAGTT
80	orf111a.pep	651	GCACCGGAA	GNCAAAACAG	CGGGGGCGGA	ACCTTGCCCC	ATCGGCATCG
	orf111	701	AACACGCCAA	CATCGTCCTA	GGGGCAATA	CGCAGATTAT	CGTCCCCGTC
85	orf111a.pep	751	AACACCCCTT	CNTTGCCCA	TTCCGGGGCAT	TACCGTATT	TCCACGCTGA
	orf111	801	TAAAGACGCG	AAACGCTCT	CCCATATCAT	TAATCGCAAC	AAACAAACGAC
90	orf111a.pep	851	CCATCAGCGA	AAACCTCGCC	TCCATCAGCG	TGNTCCGAGA	CAGTCGGATG
	orf111	901	ACGGCGGAGC	GCTTNTCCGAC	AGGA/TATTC	TGATTTGGCC	AAACCGAACG
95	orf111a.pep	951	CTTAAAGCTG	CGAGAGGGGG	AAAAACTCGC	TGTTTTCTG	ATTCGTCAGGG
	orf111	1001	ATAAAGCGG	CTACCGCACC	GGCATGTCTT	CGGAATTGAA	AAACATGCTC
100	orf111a.pep	1051	CGCTAA				

The complete length ORF111a nucleotide sequence <SEQ ID 447> is:

40	1	ATGCCGTCG	AAACACGCTC	GGCGAACCTT	ATCCGCACCT	TGATATTG	CG
51	CCTGGATTTT	ATCTCTCTGA	AGCGCTGTC	GGAAACAAAC	GGCCAAACCG		
101	TTACCTCTGA	AGGTGAAAGC	ATGSGCAGA	CCTATAACCT	CAAATACCTT		
151	TCAATAATC	GGGACNACT	CCCTCACT	GGCGAAATAC	AAAANGCAT		
201	CGATGACCGG	CTTAAAGAAG	TCAACCGGG	GATCTCCACC	TATCAGCCCG		
251	ACTCGAAT	CAGCGCGTTC	AACCAACRA	CAGCGCGCMA	GGCCCTCCGG		
301	ATTCAACGCG	ACTTGCGACA	CGTTACTCGC	GAAGCGCTG	ACCTGAACCG		
351	CCGTGACACAC	GGCGCGCTG	ACGTAACCGT	GGGCCCCCTTG	GTCAACCTTT		
401	GGGGATTTGCG	CCCCGACAAA	TCGGTATCCO	GTGACCGTC	GGCGAACAA		
451	ATCAAACAAAG	CAGCATCTTA	TACGGCCATA	GACAAATATCA	TTTGAAACAA		
501	AGGCAAAAGAT	TACCGTTCT	TGAGCAAAAC	CCACCCCAAG	GCCTATTG		
551	ATTATATCTTC	GATGGCCAA	GGCTTCGGCG	TTGATNANT	TGGGGCGCAA		
601	CTGGAAAAAT	ACGGCATCTA	AAATATTCAT	GTGCAAATCG	GGCGNGAGTT		
651	GCACCGGAA	GNCAAAACAG	CGGGGGCGGA	ACCTTGCCCC	ATCGGCATCG		
701	AACACGCCAA	CATCGTCCTA	GGGGCAATA	CGCAGATTAT	CGTCCCCGTC		
751	AACACCCCTT	CNTTGCCCA	TTCCGGGGCAT	TACCGTATT	TCCACGCTGA		
801	TAAAGACGCG	AAACGCTCT	CCCATATCAT	TAATCGCAAC	AAACAAACGAC		
851	CCATCAGCGA	AAACCTCGCC	TCCATCAGCG	TGNTCCGAGA	CAGTCGGATG		
901	ACGGCGGAGC	GCTTNTCCGAC	AGGA/TATTC	TGATTTGGCC	AAACCGAACG		
951	CTTAAAGCTG	CGAGAGGGGG	AAAAACTCGC	TGTTTTCTG	ATTCGTCAGGG		
1001	ATAAAGCGG	CTACCGCACC	GGCATGTCTT	CGGAATTGAA	AAACATGCTC		
1051	CGCTAA						

This encodes a protein having amino acid sequence <SEQ ID 448>:

```

1 MPSETRLPNF IRTLIFALSFIFI
5 SNNRDXLPS AEIQKRIDDALEVNRMQSTYQ
10 PDSEISRFNQHTAGKPLRISSDFAHVTAEAVRLNR
15 LTH
20 GALDTVTVGPLVNLWGF
25 GGDVKVAGELEKYG
30 IYLV
35 EIGGELHGKX
40 KNARGEPRIGIEQPNI
45 VQ
50 SAM
55 DSAM
60 VVADSAM
65
70
75
80
85
90
95
100
105

```

101 ISSDEFAHVTAAEVHLNRLTH GALDVTVGPL VNLIWGFGPDK SVTREPSPEQ
 151 IKQAASYTGI DKIIILKQGKD YASLSKTHPK AYLDLSSIAK GFGVDXVAGE
 201 LEKYQIQLY VELIGEGLHGK XRNARGEFWR IGIEQPNIVQ GGNTQIIVPL
 251 NNRSXATSGD YRITHVDKSG KRLSHIINPN NKRPISHNL A SISVXADSAM
 5 301 TADGXSTGLF VLGETEALKL AEREKLAVFL IVRDKGGYRT AMSSEFEKLL
 351 R*

Homology with a predicted ORF from *N.gonorrhoeae*

ORF111 shows 96.6% identity over a 351aa overlap with a predicted ORF (ORF111.ng) from *N.*

10 *gonorrhoeae*:

		10	20	30	40	50	60
	orf111ng	MPSETRPLPNLIRALI	FALGFIFLNA	CSEQTAQTVTLQGETM	GTTTVKVLYSNNRDKLPSP		
15	orf111	MPSETRPLPNFIRVL	I	LAFALGFIFLNA	CSEQTAQTVTLQGETM	GTTTVKVLYSNNRDKLPSP	
		10	20	30	40	50	60
	orf111	70	80	90	100	110	120
20	orf111	AKIQKRIDDALKEVNRM	QMSTYQD	SEISRFNQHTAGKPLR	I	ISSDFAHVTAAEVHLNRLTH	
		70	80	90	100	110	120
	orf111ng	130	140	150	160	170	180
25	orf111	GALDVTVGPLVNWLW	GFGFPDKSVT	REPSPPEQIKQAA	ASYTGIDKII	LQGKDYASLSKTHPK	
		130	140	150	160	170	180
30	orf111ng	AYLDSLSSIAKGF	GFDKVAGELE	KYQIYLN	VIEIGELHKGKGN	GHGEPWRI	IGEQPNIIQ
	orf111	AYLDSLSSIAKGF	GFDKVAGELE	KYQIYLN	VIEIGELHKGKGN	GHGEPWRI	IGIEQPNI
		190	200	210	220	230	240
35	orf111ng	GGNTQIIPN	PLNNRSLATS	GDYRIFHV	DKGRLSHI	IINPNNKRPI	SHNLASISVVSDSAM
	orf111	GGNTQIIPN	PLNNRSLATS	GDYRIFHV	DKGRLSHI	IINPNNKRPI	SHNLASISVVADSAM
		250	260	270	280	290	300
40	orf111ng	TADGLSTGLF	VLGETEALKL	AQEKLAVFL	IVRDKG	YRTANSSEFA	KLRLX
	orf111	TADGLSTGLF	VLGETEALKL	AQEKLAVFL	IVRDKG	YRTANSSEFE	KLRLX
45		310	320	330	340	350	

The complete length ORF111ng nucleotide sequence <SEQ ID 449> is:

1 ATGCCGTCTG AAACACGCCG GCGGAACCTT ATCCGCGCTT TGATATTGCG
 51 CCTGGTTTC ATCTTCTGA ACSCCTGTC Gaaacaacac GCGCAaaccg
 101 TTACCTCTGA AGCCGAAAGc atGGGTACGA CCTATACCTG CAAATACCTT
 151 TCAAATACT CGGCACAACT CCCCTCCCCC GCCAATACAT AAAAGCCAT
 201 TGATGATGCG CTAAAGAAG TCACCCGGCA GATGTCACCC TACAGACCG
 251 ATTCCGAAAT CAGCGGGTTC AACCCACACA CAGCGGCCAA GCCCCCTCCG
 301 ATTCGAAGCG ATTCGGCAGC CGTACCGGCC GAAGCCGTCG GCTGAACCG
 351 CCTGACTCAC GCGCCACTTG ACCTAACCGT CGGCCCTTG GTCAACCTT
 401 GGSGGTTCTGG CCCCGACAAA TCCGTTACCGA GTGACCGTCG GCGGAACCAA
 451 ATCAAACAGG CGGCATCTT TACGGGCATA GACAAAATCA TTTTGCAACA
 501 AGGCGAAAGAT TACCTCTCT TGAGCAAAAC CCACCCCCAA GCTCTATTG
 551 ATTATCTTC GATGCCAA GGCCTCGCG TTGATAAAGT TGCGGGCCAA
 601 CTGGAAAAAT CGGCATCA AAAATCTCG GTCGAAATcg gcggGAGTT
 65 651 CGACCGCCAA GCGAAAAATG CGACCGGGCA ACCCTGGCGC ATCGGTATAG
 701 AGCAACCCAA TATCATCAGA GgcgGCAata CGCAGATTat cgtccgcgt
 751 aacCaacggc cgctTGCCAC TTCCGGCGAT TaccgtatTT tccacgtcgA
 801 TAAAACcgcc aaacggccccc acatCCCAca acAAACGac
 851 ccATCGCaca caacccgtccccc tccatcagcg tggtctcAGA CAGTGCAGT
 901 AGCGCGACG GTTtatCAGA AGGATATTGTTTGGCG AAAACCGAAGC
 951 CTTAAGGCTG CGAGAACAAAG AAAACCTCGC TGTTTTCTPA ATTGTCGCG

1001 ATAAGGACGG CTACCGCACC GGCATGTCTT CGGAATTGTC CAAGCTGCTC
 1051 CGCTAA

This encodes a protein having amino acid sequence <SEQ ID 450>:

5 1 MPSETRPLNL IRALIFALGF IFLNACSEQT AQTVTIQLGET MGTTTVVKYL
 51 SNNRDKLPSF AKIQKRRIDA LKEVNRMQST YQDTSIEISR P NQHTAGKPLR
 1001 ISSDFAVHTA EAVRLNRLTH GALDVTVGPL VNLWGFGPDK SVTREPSPSEQ
 151 IKQAASYTGI DKIIILQDGKX YASLSKTHPK AYLDLSIIAK GFGVVDKVAGE
 201 LEKYQIQNLY VEIGGELHNGK GKNABGEPEWIG IEQEPNIIQ GGNQTQIVPL
 251 NNRSLATSGD YRITHVDDKNG KRLSHIINPN NKRFISHNLA SISUVSDSAM
 301 TADGLSTGLF VLGEETEALRL AEQEKLAVFL IVRDKDGYRT AMSSEPAKLL
 351 R*

This protein shows homology with a hypothetical lipoprotein precursor from *H.influenzae*:

15 sp|P44550|YOJL_HAEIN HYPOTHETICAL LIPOPROTEIN HI0172 PRECURSOR >gi|1074292|pir|4
 hypothetical protein HI0172 - *Haemophilus influenzae* (strain Rd W20)
 >gi|1573128 (U32702) hypothetical [Haemophilus influenzae] Length = 346
 Score = 353 bits (89%), Expect = 9e-57
 Identities = 181/344 (52%), Positives = 247/344 (71%), Gaps = 4/344 (1%)
 20 Query: 7 LPNLIRALIFALGF IFLNACSEQT AQTVTIQLGET MGTTTVVKYL
 + LI +I + L AC ++T ++L G-TMGTYY VKYL + S K +
 Sbjct: 1 MKKLISGIIAVAMSLAACQKET-KVISLSGKTMTGTTYHVVKYLDDGSIATSE-KTHEE 58
 25 Query: 67 IIDALKEVNRMQSTYQDSEISRFNQHT-AGKPLRISSDFAHVTAEVRLNRILTHGALDV 125
 + LK+VN +ASTY+ DSE-SRFNQ+T P+ IS+DFA V AEA-RIN++T GALDV
 Sbjct: 59 IEAIKLDVNAMKSYTKDSELSRFNQNTQVNTPIEISADFAVKLAIRLNKVTGALDV 118
 30 Query: 126 TVGPVLNVLWGLGPDKSVTRESPQEQIKQAASYTGDIIK1LQQGKDYASLSKTHPKAYLDL 185
 TVGP+VNLWGLGP+K ++P+FEQ+ ++ GIDK1 L K+ A+LSK P+ Y+DL
 Sbjct: 119 TVGPVLNVLWGLGPDKSVTRESPQEQIKQAASYTGDIIK1LQQGKDYASLSKTHPKAYLDL 178
 35 Query: 186 SSIAKGFGVDRKVAELEYKGQNYLVEIGGELHNGKRNNAHGEPEWRIGIEQPNIIQGGNTQ 245
 SSIAKGFGV+D+V+ +LE+ QNY+VEIGG+ EKGNK G+FW+I IE+P +
 Sbjct: 179 SSIAKGFGVDRKVAELEYKGQNYLVEIGGELHNGKRNNAHGEPEWRIGIEQPNIIQGGNTQ 238
 40 Query: 246 IIVPLNNSRLATSGDYRIPHVWDKNGKRLSHINPENNKRPISHNLASIVSWSAMTAGDL 305
 + LNN +A+SGDYR+ ++NGKR +H I+P PI H+LASI+V++ ++NTADGL
 Sbjct: 239 AVIGLNGKMASSGHDYR+FEENGKRAFHEEDPPTGKYPIQHHLASITVLPSTSMTADGL 297
 Query: 306 STGLFVLGEGDRAEQLRLEQOKLAVELIIVRDKDGYRTAMSSEPAKL 349
 STGLFVLGEGDRAEQLRLEQOKLAVELIIVRDKDGYRTAMSSEPAKL 349
 Sbjct: 298 STGLFVLGEGDRAEQLRLEQOKLAVELIIVRDKDGYRTAMSSEPAKL 341

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 54

45 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 451>:

5 1 ..CCCTGCCGCC GACAGGGCGA CGACGCTGTAT GGGGCCGAG CGTCCCGTCA
 51 AAAATCTGG CTGGCGCTCA TCGGCCGCGC GTGCCATCAA AATATAACGG
 101 GCGCGCGCGC TCGCGGAGG TGCGCGAAG GGGTGAACAAT CGCGCGCGAG
 151 GTGTTTGATC GGCGAAANTGA AGGCCACCKA YTGGCAACTCG GCGTGATGCG
 201 CGCGCAGGCC GGCCCGAGC CWTCACTCAA CGCGAAAGCGC GTGCGCGCAG
 251 gCATGATGTTA GTATGTTGATG TTGGGGGTG TTTATCTGTC GTGGCATCAG
 301 TTGGCGGATA AACAAACGGG TGCTTATTC GACCGCTGTG TGCATATACCA
 351 ACGTTTCAAAC CACCGGGG ATGATGAAAGG CGGTGCGGAA CGCTAACAAA
 401 CCAAGGTTG GACGGCTCT GTGGAAGGG CCTACACCGC GCTTGTGGCG
 451 GAAAGCATTTG TCGGAAGGAGG CAAATATGTC CGGTTTTTACCC TACAACCGCA
 501 GgcGCGAGTTT ACCACTTGG GCCTTAACGG CGCGTTTACCC GACAGCGAGG
 551 CGACGGCGGT CGGAGCTGCTC GGCAAGGGCTC AGTGGCAAGG CGCGCGCGGC
 601 AtTCGGGCAA AAAACCGCTT TGCTTGTGGT AACGGTGTCA ATCTTCAGCC
 651 TTTTGGCGCT TTTAATGTTT TGCACAGGTC AAAATCTTC GGCGTGGAAA
 701 TGCGACGGCGA AAAACAGAGC CTGGCAGGCA GGACGGCACT CGAAGGGCGG

751 TTCCGGTATTG AAGCCGGTTG GAAAGGCCAT ATGTCGGCA..

This corresponds to the amino acid sequence <SEQ ID 452; ORF35>:

5 1 .. PCRRQGDDVV AAHASRQLW LRFIGGRSHQ NIRGAAAADG WRKGVQIGGE
 51 VFFVRQNEGSGX LAIGVMGGRG GQHASVNGKG GAAGSDLYGY GGGYAAWHQ
 101 LRDKQTGAYL DGWLQYQRFH HRINDENRAE RYKTKGWTS VEGGYNALVA
 151 EGIVWKGNNV RFYLQPOQAF TYLGVNGGFT DSEGTAVGLL GSQGWQSRA
 201 IRAKTRFLALR NGVNLQPFFA FNVLHRSKSF GVEMDGEKQT LAGRTALEGR
 251 FGIEAGWKGH MSA..

Computer analysis of this amino acid sequence gave the following results:

10 Homology with putative secreted VirG-homologue of *N. meningitidis* (accession number A32247)

ORF and virg-h protein show 51% aa identity in 261aa overlap:

15 Orf35 5 QGDVYVAAHSRQLWLRFIGGRSHQNIRGAA-ADGWRKGVOIGGEVFRQNCGSXLAT 63
 + D++ R+ LWL 1 G S+Q ++G A +G+RKGVQ+GGEVF QNE + L+I
 virg-h 396 KNSIDFDTLPRKGWLWLRVIDGHNSQWVQKGTAPEVGSYRKGVQLGGEVFTWQNESNLQSLI 455
 Orf35 64 GVMGRAGQHASVNGKG--GAAGSDLYGYGGVYAAWHQLRDKQTGAYLDGWLQYQRFH 121
 G+MGG+ Q ++ + ++ G+ GVTVA WHQI+DKTGAYD W+YQRFH
 virg-h 456 GLMGQQAECRSTHNPDTDNLTGNGVKGFGAGVYATWHQLDKQTGAYDWSWMQYQRFH 515
 20 Orf35 122 RINDENRAYERKTKGWTAASVEGGYNALVAGT VKGNNVRFYLQPQAOQFTYLGVNGGFTD 181
 RIN E+ ER+ +KG TAS-E GYNAL-AE KGN+R YLQPQAOQ TYLGVNG F+D
 virg-h 516 RINTTEDGTERFTSKGITAISIEAGYNAALLAEH+TTKGNNSLRVYLPQAOQLTYLGVNGKFSD 575
 25 Orf35 182 SECTAVGLLGSQWQSRAIGTRAKTRFALRNQVNLQPFPAAFNVLIRSKSFGVEMDGEKQTL 241
 SE + LLGS Q C+R G++AK F+F-L + +PPFAA N L+ +K FGVEMDGE++ +
 virg-h 576 SENAHVNLLGSRQLCQTRGVQAKAQFSLYKNIAIEPPFAAVNALYHNKPFVGVEMDERRVI 635
 Orf35 242 AGTAGLEGRGIEAGWKHMS 262
 +TAE+ +G+ K H++
 virg-h 636 NNKTAIESQLGVAVKIKSHLT 656

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF35 shows 96.9% identity over a 259aa overlap with an ORF (ORF35a) from strain A of *N. meningitidis*:

		10	20	30
35	orf35.pep	PCRRQGDDVVAAHASRQLWLRFIGGRSHQNIRG		
		:		
40	orf35a	QRЛАІРЕАВЛІЯАQQАYАANTLFGILRADRGDDVYAADPSRQKLWLRFIGGRSHQNIRG	310 320 330 340 350 360	
45	orf35.pep	GAAADGWRKGVOIGGEVFRQNCGSXLAIGVGMGRAGQHASVNGKGAAAGSDLYGYGGGV	40 50 60 70 80 90	
	orf35a	GAAADGRRKGVYQIGGEVFRQNCGSRLAIGVGMGRAGQHASVNGKGAAAGSYLHGGGV	370 380 390 400 410 420	
50	orf35.pep	YAAWHQLRDKQTGAYLDGWLQYQRFKHRINDENRAYERKTKGWTAASVEGGYNALVAEGIV	100 110 120 130 140 150	
	orf35a	YAAWHQLRDKQTGAYLDGWLQYQRFKHRINDENRAYERKTKGWTAASVEGGYNALVAEGVV	430 440 450 460 470 480	
55	orf35.pep	GKGNVNRFYLQPQAOQFTYLGVNGGFTDSEGTAVGLLGSQWQSRAIGIRAKTRFLRNQVN	160 170 180 190 200 210	
	orf35a	GKGNVNRFYLQPQAOQFTYLGVNGGFTDSEGTAVGLLGSQWQSRAIGIRAKTRFLRNQVN	490 500 510 520 530 540	
60	orf35.pep	LQPFAAFNVLRSKSFVGVEMDGEKQTLAGRTALEGRFGIEAGWKGHMSA	220 230 240 250 260	

-272-

	orf35a	LQPFAAFNVNLHRSKSFGVEMDGEKQTLAGRFRGIEAGWKGHMSARIYGKRTDGD 550 560 570 580 590 600
5	orf35a	KEAALSLKLWLFX 610 620

The complete length ORF35a nucleotide sequence <SEQ ID 453> is:

1	ATGTTTCAGAG	CTCGACTTGG	TTCAAAATACT	CGTTCTACCA	AAATCGGCGA
51	CATGTCGGAT	TTTCATCAG	CAGACAAAGCC	GAIAACCCGGC	ACTTCCTATT
101	ATTTTTCCAG	CGGTAAAACC	GATCAAAATT	CATCCGAAATA	TGGGTATGAC
151	GAATACATA	TCCRAAGTTAA	AAACTACAAAT	ACCGGCATAC	TGCCCGTCGA
201	TAATATGCC	GGTTGAAAGA	AAATTTTACAC	AGAAACTTAC	GGGGATTAATT
251	TAAGGGATGC	GGTTGAAAGA	CAATTACAGG	ATTATACAA	AACAAGACCC
301	GAAGCTTGGG	AAAAGAAAGA	AAAACGGACT	GAGGGAGGCGT	ATATAGAACAA
351	GCTTGGACCA	AAATTAGTGA	TACTAACAA	AAAAACCCC	GATTTAAATA
401	ATAAATTTGT	AGGAAGATTC	GTACTCACTC	CTCATAGTAA	TACATCACAG
451	ACTAGTCTCA	ACAAACATCT	CAATAAAAAA	TTACACGTC	AAATCGAAAA
501	CAAAATCCAC	TGCGCCGNC	AGGTGTTGGA	ACTGACCAAAG	ATGACGCTGA
551	AAAGATTCCT	TTGGGACGCC	CGCCGCCATT	CGGACATCCA	TAATGCTGGA
601	ACTTCCGATA	ATGCCCGCT	CCCGCTTGAC	ACCGAAAGATG	AAAACATGAC
651	CGTCCTCATAAA	GGCTGATCAG	CGCGTGCAGG	TTTCTCTTC	GGCTGACAGC
701	TGCGGTGGAGTC	GGAAACACCC	GGCCCTGACCT	TTGAGAAAGAA	ATGTCAGCGGA
751	CAATCCGGCG	TGGTTTTGG	ACGCCGGCC	AAAAATCTGA	AAACGCTGGA
801	CGGGCGCAAA	TCTATTCGCG	CGGAAAAGGC	AGACTCTTAAT	TCTTTTGGGT
851	TTAAACAAA	TTACCCGGC	GGACTGTACG	AATTATGCT	CAAGCAATGC
901	GAAGGCGGAT	TTTCGCTTGG	CGTGCAGCGT	TTGGCTATCC	CCGAGGGCGGA
951	AGGGCTTTTA	TATGCCCAAC	AGGCTTATGC	GGCAAATACT	TTGTTGCGGC
1001	TGCGGTGGCG	CGACAGGGGC	GAAGCAGTGT	ATGCCGCGGA	TCCGTCGCCGT
1051	CAAAATTTGT	GGTCGCGCT	CATCGCGGCC	CGGTGCGCATC	AAATAATACG
1101	GGGGCGGGCG	GCTCGCGGC	GGACGCGCAGA	AGGGGTGCHA	ATCGCGGGCG
1151	AGGTGTTTGT	AGCGCAAAAT	GAAGGCGACCC	GGCTGGCAAT	CGCGTGTATG
1201	GGGGCGAGGG	CTGGCGACAG	CGGATCAGTC	AAACGCAAG	CGCGTGCAGC
1251	AGGCAGTTAT	TTGCATGTT	ATGGCGGGGG	TGTTTATGCT	CGCTGGCATC
1301	AGTTCGGCA	TAACAAACAG	GGTGGCTATT	GGCGCGCTG	TTGCAATAC
1351	CAACATTCTCA	ACACCGCAT	CAATGATGAA	AAACCGTCCG	AACCGTACAA
1401	AAACCAAGGT	TGGACGGCT	CTGTCGAGG	CGGCTCACAA	GGCCTTGTGG
1451	CGGAAGGGCT	TGGCGAAAAA	GGCAATAATG	TGCGGTTTTA	CTCGCAACCG
1501	CAGGGCGCAT	TTTACACTT	GGCCGTTAAC	GGCGGCTTTA	CCACAGCGGA
1551	GGGGACGGCG	TTCTGGACTC	TCGGCAGCGC	TCAGTGGCCA	AGCCGCGGCC
1601	GCATTCTGGC	AAAACCCAG	TTTGCCTTTC	GTAAACGTTG	CAATCTTCAG
1651	CCTTTTCCG	CTTTTAATGT	TTGCAACAGG	TCAAAATCT	TCCGGGTGGA
1701	ATTGGACCGC	GGAAACACCA	CGCTGGCAGG	CAGGACGGG	CTCGAAGGGC
1751	GGTTCGGCAT	TGAAGCCGGT	TGAAAAGGCC	ATATGTCGCG	ACCCATCGGA
1801	TACGGCAAAA	GGACGACCGG	CGACAAAGAA	GGCGCATGTG	CGCTCAATG
1851	GCTGTTTGA				

45 This encodes a protein having amino acid sequence <SEQ ID 454>:

1	MFRQALGSNT	RSTKIGDDAD	FSFSDKPKPG	TSHYFSGGKT	DQNSSEYGYD
51	EINIOGKNYH	SGLAVDNNA	VVKVKTIDTY	DGNILKDAVKK	QJODLYKTRP
101	EAEWENKKRT	EEAYIEQLGP	KFSILKQKNN	DLINKLVEDS	VITPHSNTSQ
151	TSLNIRNFKN	LHVKVIRLN	TKDEKLTVHK	AYQGGADEFL	GVDRESDKP
201	TSNDARIRLN	TKDEKLTVHK	AYQGGADEFL	GVDRESDKP	ALTFEEKVSG
251	OSGVLLIERP	ENLKTLDRGK	LIAEAKADSN	SAFKQNYRQ	GLYELLLKQC
301	EGGFICLVQR	LAIPEAEAVL	YAQQAAYANT	LEPLGRADRG	DDVYAADPSR
351	OKLWLRLFIGG	RSHONIRGGA	AAIDGRKGKVQ	IGGEVVFVRQN	EGSRILAIGVM
401	GGRAOGHHSV	NGKGGAAGSY	LHGYGGGYVA	AWHOLADKOT	GAYLQDWLOY
451	QRFKIRINDE	NRAERYKTG	WTASVEGGGVN	ALVAEGVVGK	GNNNVRYLQP
501	QAOFTYLGVN	GGFTDSEGT	VGLLGSGQWO	SRAGIRAKTR	FAIRNGVNLO
551	PFPAFNVLHR	SKSFVGEMDG	EKQTLAGRTA	LEGRFRGIEAG	WKGHMSARIG
601	YGRKTDGKE	AALSKLWLF	*		

Homology with a predicted ORF from *N. gonorrhoeae*

60 ORF35 shows 51.7% identity over a 261aa overlap with a predicted ORF (ORF35ngh) from *N. gonorrhoeae*:

orf35.pep	PCCRQGDDVVYAHASRQLWLRFIGGRSHQNIRG	34
orf35ngh	::::: :: : :::: :: :	370

	orf35.pep	GAA-ADGWRKGVQIGGEVFWRQNEXSLAIGVMGRAGQHASVNGK--GAAGSDLYYG	91
	orf35ngh	: : : : : : : : : : : : : : : : : :	
5	orf35.pep	KTAPVEGYRKGVQLGGEVFTWQNESNQLSIGLMLGGQAEQRSTFRNPDTDNLTGNVKGFG	430
	orf35ngh	KTAPVEGYRKGVQLGGEVFTWQNESNQLSIGLMLGGQAEQRSTFRNPDTDNLTGNVKGFG	
	orf35.pep	GGVYAAWHQLRDKQTGAYLDGWLDYQRFCHRINDENRRAERYTKXGWTASVEGGYNALVAE	151
	orf35ngh	: : : : : : : : : : : : : : : : : : : : : : : : : : : : :	
10	orf35.pep	AGVYATWHLQDKQTGAYDWSMMQYQRFRHRINTEYATERFTSKGITASIEAGYNALLAE	490
	orf35ngh	AGVYATWHLQDKQTGAYDWSMMQYQRFRHRINTEYATERFTSKGITASIEAGYNALLAE	
	orf35.pep	GIVGKGNVRFYLPQQAQFTYLGVNGGFTDSEGTAVGLGSGQWQSRRAGIRAKTRFALRN	211
	orf35ngh	: : : : : : : : : : : : : : : : : : : : : : : :	
15	orf35.pep	HTTCKGNLSRLVYLPQPAQTLTYLGVNGKFDSENAQVNLLGSRQLQSRVGVQAKAQFAFTN	550
	orf35ngh	HTTCKGNLSRLVYLPQPAQTLTYLGVNGKFDSENAQVNLLGSRQLQSRVGVQAKAQFAFTN	
	orf35.pep	GWNLQPFAAVNLHRSKSFGEVMDGEKQTLAGRTALEGRFGIEAGWKGHMSA	263
	orf35ngh	: : : : : : : : : : : : : : : : : : : : : : : : :	
	orf35ngh	GVTFPFVVAVNSTQQKPFGEIDGDRVNNKTVIETQLGVAAKIKSHLTLQASFNRQT	610

A partial ORF35ngh nucleotide sequence <SEQ ID 455> is predicted to encode a protein having partial amino acid sequence <SEQ ID 456>:

20	1 ..KKLDRRNSEY WKEETYHIKS NGRTYPNIPA LFPKHPFDFF ENINNSKKIS
	51 FYDKEYTEDY LVGFARGFGV EKRNGEEKEP LRQYFKDCVN TENSNNNDCK
	101 ISSFGNYGPI LIKSDIFAL A SQKNSHNS EILSVGNYIE WLRPTLNKL
	151 GWOQHLYAGL DFFHYIEVTD NSHVIGQTID LGALELTNSL WKPRWNNSID
	201 YLTIKNAEIR FNTKNESSLV KEDYAGARGF RFAYDLKDKV PETPYVLT
	251 NUTGTSDII ECGKALDLKH LDHQIYKVNV DTADKDKDFL SSKYRKKGYYT
	301 LSLQQRPEGE FTWKVERDOI AIIAQQAQAA NTFLPARLND KNSDIFDRTL
	351 PREGGLWLRI DHGSNQWVQG KTAPVEGYR GVQLGEGEVT WQNESNQLSI
	401 GLMGQAAEQ SFTFRPDTDN LTITGNVKGGF AGVYATWHLQ DKRQTGAVZD
	451 SWMVOYRFRU RINTYEYATERFTSKGITASIEAGYNALLAE HTTCKGNSLR
	501 VYLQVPAQLT YLGVNKGFD SDENAQVNLLG SRQLQSRVGV QAKAQFAFTN
30	551 GVTFPFVVAV NSTQQKPFGEIDGDRVNNKTVIETQLGVAAKIKSHLTLQASFNRQT
	601 TLQASFNRQT SKHHHAKQGA LNLQNTF*

Based on this prediction, these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 55

35 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 457>:

40	1 ..GGCGAATATG TTCAGTTCTC TATAGATTTC TTCAAGTGTGG GTAATACGGG
	51 GGGCGGTATA CCTAACGCTA ACCGCTGTGTT TGATGCGAAA CGGAGATGGG
	101 AGGTGATAGA GAACTCTTAAT AAATGACAA CTCCGTGAGCA GTTGGAGAAA
	151 AATGTTCAAGG AACAGCAGAG AAGGGACTGAG ACTAGTCAGT TAAAGGCCA
	201 TGCCGCAACGG ATAATGGAAA ATAACACAGG GTTAGATTTT AATCATTTTA
	251 TAGGTGTGTA TATCAATAAA AAAGGCACAG TRACAGGAGG GCATAGCTTA
	301 ACCCGTGGTG ATGTAACGGT GATACACCAA ACCTCGGCAC CTGATAAACAA
	351 TGCGGT..TTA TCAAGCGACA GTGGAATTT A

This corresponds to the amino acid sequence <SEQ ID 458; ORF46>:

45	1 ..AEYVQFSIDL FSVGKSSGGI PKAKPVDAK PRWEVDRKLN KLTTREQVEK
	51 NVQETRRSQ SSQFKAHQQR EWENKTGLDF NHFIGGDINK KGTVTGGHSL
	101 TRGDRVRVIIQ TSADPKHGXLS SDDSGNX

Further work revealed further partial nucleotide sequence <SEQ ID 459>:

50	1 ..GCAGTGTGGC TnCCGATGCGA TGCAACGCC CTCAnATTGGA CAAACGATTG
	51 TTTTATCCGG CAGGTCTCTCG ACCGTCAGCA TTTGGCAACCC GACGGGAAAT
	101 ACCACCTATT CCCACCAAGG CGCGAACCTTG CGCACCCCCA CTCTCATATC
	151 GGATTGGAA AAATACAAAG CCATCAGTGT GGCAACCTGA TGTTCGAACA
	201 GGCGCCATTAA AGAACGAAATA TCGCTCATGT TGGCGCTTGT TCCGATCACG
	251 GGCAAGAACG CTATTCCTCS TTGGCAACACC ATGGCTCACA TTCCGATCTT
	301 GTGAAGGCG TGATGCCCGT TGACGGATT AGCCTTFAAC GCATCCATTG
	351 GGACGATAC GAACACCATC CGCGCAGGG ATATATACAG TTACGACATA
	401 CGCGCTATCC CGCTCCAAAAA GGCGCGAGGG

-274-

451	AAAGGCGTTG	CCCAAAATAT	CCGCCTCAAC	CTGACCGACA	ACCGCAGCAC
501	CGGACAAAGG	CTTCGGCACC	GTTTCCACAA	TGCCGGTAGT	ATGCTGACGC
551	AAAGGAGTAGG	CGACGGATTG	AAACGGGCCA	CCCGATACAG	CCCGAGACTG
601	GACAGATCGG	GCCTATGCCG	CGAACGCTTC	AACGGCAGTG	CAGATATCGT
651	TAACAAACATC	ATCGGGCCTG	CAGGGAART	TGT	

5 This corresponds to the amino acid sequence <SEQ ID 460; ORF46-1>:

1	..AVCLPMPMHAA	SXLANDSFIR	QVLDRQHFEV	DGKYHLFGSR	GELAERQSHSI
51	GLGKIQSHQL	GNLMLIQQAL	KGNIGYIVRF	SDBHGHEVHSR	FUNHASHSDS
101	DEAGSPVWGF	SLYRHWWDY	EHHPADGYDG	POGGGGYPAPK	GARDIYSYDI
151	KGVANQNLRLN	LTDNRSTGQR	LADRFHNAGS	MLTQGVGDGF	KRATRYSPEL
201	DRSGNAAEAF	NSTADIVKNI	IGAAEIL		

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.gonorrhoeae*

ORF46 shows 98.2% identity over a 111aa overlap with a predicted ORF (ORF46ng) from *N.*

15 *gonorrhoeae*:

orf46.pep	AEYVQFSIDLFSVGKGSSGGIPKAKPVPFDKPRWEVDRKLNLKLTTR	45
orf46ng	PKTGVVPFDKGKGPFPNFKEHVKYDTKLQELQELSGGGIPKAKPVPFDKPRWEVDRKLNLKLTTR	217
orf46.pep	EQVEKVNQETRRRSQSOFKAHAQREWEENKTGLDFNHFIGGDINKKGTVTGGHSLTRGDV	105
orf46ng	EQVEKVNQETRRRSQSOFKAHAQREWEENKTGLDFNHFIGGDINKKGAUTGGHSLTRGDV	277
orf46.pep	RVIQQTTSAPDKHGXLLSSDSGN	126
orf46ng	RVIQQTTSAPDKHGVLLSSDSGN	298

25 A partial ORF46ng nucleotide sequence <SEQ ID 461> is predicted to encode a protein having partial amino acid sequence <SEQ ID 462>:

1	..RRLKHCCCHAR	LGSAHPRHKQD	GAHQCRPGRYR	ATQRLCRSSH	PRLGSPKPQC
51	RTRHRSRQQY	LYGSHSPHQRD	WSPGKJQILGQ	RHHTGTSRAV	ADXDRDRICER
101	RTRHRSRQQY	CRLGKPSLSS	IPKYPPLKLEQ	RYGKENITSS	TVPPSNKGKV
151	RЛАDQRHQFKT	GVPFGGGGF	NFERHVKYDT	KLDIQLESGG	G1PKRKPVFD
201	AKPRWEVDRKLNLKLTTR	EQEVNQETRRR	SQSOFKAHAQREWEENKTGL	DFHNFIGGDI	NKGVAVTGGH
251	DFHNFIGGDI	NKGVAVTGGH	SLTRGUVVRV	QQTSA	PDKHGVLLSSDSGN*

35 Further work revealed the complete gonococcal DNA sequence <SEQ ID 463>:

1	TTGGGCATTTC	CCCGCAAAAT	ATCCCTTATT	CTGTCCCATAC	TGGCAGTGTG
51	CTCGGGATG	CTTCGACATG	CTTCGACATTG	GccaaACGAT	CCTTTTATCC
101	GcCagggttc	TGaccggTCAG	CATTTCGaac	ccgacggGta	ATACCACTTA
151	Tgcggcga	CA	GGGGGGAGCT	TGccnagcGC	aacggccATA
201	aaacatAcaa	Agccatca	cgat	tgggactt	tccgattttca
251	ttagaaagaa	TATcgGct	ttatggccat	tttccgatca	cggcacaaaa
301	ttccatcg	cccccGca	ccatGCTCTA	CATTGCGATT	CTGACGAAAC
351	CGGTAGTCCC	GTTGAGCGGT	TCAGCCTTTA	CCGGCATCCAT	TGGGACGGAT
401	ACGAACACCA	TCCCGCCGAC	GGCTATGACG	GGCCACAGGG	GGCCGCGTAT
451	CCGGCTCCCA	AAAGCCGGAG	GGATATATAC	AGCTACGACA	AAAAGGGT
501	TGCCCCAAAT	ATCCGCTCA	ACCTGACCGA	CAACCGCAGC	ACCGGACAC
551	GGCTTGCCTG	CGGTTCCTAC	AAATGCGGCG	CTATGCTGAC	GCAAGGAGTA
601	GGCGACGGAT	TCAAACGCGC	ACCCGATAC	AGCCCCGAGC	TGGACAGATC
651	GGCGCAATG	ggccGAAGCT	TCAAACGCGAC	TSCGATATAC	GTCAAAAACA
701	TCTATGGCGC	GGCAGGAGA	ATTGTGCGG	CAGGGCATGC	CGTGCagGGT
751	ATAGCGGAG	GCTCAACAT	TGCTGCTCATG	CACGGCTTGG	GTCCTGTTTC
801	CACGGAAAC	AAAGTGGCGC	GCATCAACCG	TTTGGCAGAT	ATGGGCGAAC
851	TCAAGACTA	TGCGCGACGA	GCCATCGCG	ATTGGGCACT	CCAAAACCCC
901	ATAGCGGAC	ANGGCATAGA	AGCCGTCAGC	ANTATCTTA	TGGCAGGCAT
951	CCCTCATCAA	GGGGTAGGAG	TCTGCGGGGG	AAATATACGGC	TTGGCGGGCA
1001	TCAAGGGACA	TCCCTGTCAG	CGTGGCGAGA	TGGGGCGGAT	CCGATTCGCG
1051	AAAGGGAAAT	CCGGCGTCAG	CGACATTTT	GCCGATGCGG	CATACGCCAA
1101	ATACCGGTCC	CCTTACCAT	CCCGAATAT	CGGTCACRC	TTGGGAGCGC

1151	GTTACGGCAA	AGAAAACATC	ACCTCCTCAA	CCGTGCGGCC	GTCAACGGC
1201	AAAATGICA	AACCTGGAGA	CCAACGCCAC	CCGAAGACAG	GCGTACCGTT
1251	TGACGGTAA	GGTTTCGCA	ATTITGAGAA	GCACGTGAAA	TATGATACGA
1301	AGCTCGATAT	TCAAAGAATTA	TCGGGGGGCG	GTATACCTTA	GGCTAACGCT
5	1351	GTGTTTGATG	CGAACCGAG	ATGGGAGTT	GATAGGAACG
1401	GACAACTGT	GAGCAGGTGG	AGAAAATGTC	TCAGGAAACG	AGAAAGAAGGA
1451	GTCAGAGTAG	TCAGTTTAA	CCCCATGCC	AACGAGATA	GGAAAATAAA
1501	ACAGGGTTAG	ATTTTAATCA	TTTTATAGGT	GTTGATATCA	ATAAGAAAGG
10	1551	CACAGTAAC	GGAGGGATA	TTCTAACCC	TGGTGAATGA
1601	AACAAACCTC	GGCACCTGT	AAACATGGGG	TTTATCAACG	GACAGTGGAA
1651	ATTAAAAGC	CTGTAGGAA	TTGGGAGGT	AAAACGAAA	AAAGGTGGAA
1701	AGTGTGACCA	AAGCACACCA	TGTTTCCCAA	AGATTTGGGT	GAAGGCTAGAA
1751	1751	TAGGGCGTGA	AGTTACTTCG	GCTGGGAA	GTAAATTAAT
1801	ATAAATGGC	AGGGTACAAAG	TAATCTGGGT	ATTTAAATAG	AAGGATTAC
15	1851	CGAACCTAAT	AGAACACGAT	ATCCCATTTA	TGAATAG

This corresponds to the amino acid sequence <SEQ ID 464; ORF46ng-1>:

1	LGISRKISLILSILAVCLPM	HHAASDLAND	PFIRQVLDRQ	HFEFPDGKYHL
5	51	FGSRGELAXX	NHGHIGLNIQ	SHQSLHLMIQ
10	101	FHSFEDNHS	HSDSDEAGSP	VDFGSFLYRII
15	151	PAPKGARDIV	SYDIKVGVAQN	WDGYEHHHPPAD
20	201	GDGFKRATRY	SPELDRSRGN	GYPDPGQGGY
25	251	ISEGNSIIVNH	HGLGLLSTEN	KMARINDLAD
30	301	NAAQGEAIEVS	NIFMAAPLIK	MAQLKDYAAA
35	351	GKGSVSDNF	AADAYAKAIP	AIRDWAVQNP
40	401	KNVKLADQRH	PKTGVPEDGK	FYHSNRNIRSN
45	451	VFDAPKRMEV	DRKLNLKLTTR	LEQRYKGENI
50	501	TGLDFNHFPIG	GDFINKEKHVK	TSSTEVPSNG
55	551	GDIKKGKVTY	GGHSLTRGDV	YDTKLIDQEL
60	601	IKKPDKGSWEV	TKKKGGKVTM	SGGGIIPKAKP
		NKWKQTSKSG	IKIEGTFEEN	RKAYIFIYE*

30 ORF46ng-1 and ORF46-1 show 94.7% identity in 227 aa overlap:

	orf46-1.pep	10	20	30	40
		AVCLPMHHAASXLANDSFIRQVLDRQH	FEPDGKYHLFGSRGELAER		
35	orf46ng-1	LGISRKISLILSILAVCLPM	HHAASDLANDPFIRQVLDRQH	HFEFPDGKYHLFGSRGELAXR	
		10	20	30	40
	orf46-1.pep	50	60	70	80
40	orf46ng-1	QSHIGLKGIDQS	HOLGNLMIQQAAIKGNIGYIVRFS	VDHGHEVHSPPFDNHASHSDSDEAGSP	
		70	80	90	100
	orf46-1.pep	110	120	130	140
45	orf46ng-1	NDHGIGLNIQSHQSLHLMIQ	QQAAVEGNIGYIVRFS	DNHGHFKHFSPPFDNHASHSDSDEAGSP	
		70	80	90	100
	orf46-1.pep	130	140	150	160
50	orf46ng-1	VDGFSLYRIHWDGYEHHPPAD	GYPGGYPAPKGARDIY	YDIKGVAQNIRLNLTDRNS	
		130	140	150	160
	orf46-1.pep	170	180	190	200
55	orf46ng-1	TGQRLADRFHNNAGSMLT	QGVGDDGFKRATRY	SELDRSRGNAAEAFNGTADIVKNIIGAGE	
		190	200	210	220
	orf46-1.pep	I			
60	orf46ng-1	IVGAGDAVQGISEGSNIAVMHGLLSTEN	KMARINDLADMAQLKDYAAA	AIRDWAVQNP	
		250	260	270	280

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF46ng-1 shows 87.4% identity over a 486aa overlap with an ORF (ORF46a) from strain A of

N. meningitidis:

10	20	30	40	50	60
----	----	----	----	----	----

	orf46a.pep	LGISRKISLILSILAVCLPMHAHASDLANDSFIRQVLDRQHFEPDGKYHLFGSRGELAER
	orf46ng-1	LGISRKISLILSILAVCLPMHAHASDLANDPFIRQVLDRQHFEPDGKYHLFGSRGELAXR
5		10 20 30 40 50 60
	orf46a.pep	SGHIGLGNIQSHOLGNLFICQAAIKGNIGYIVRFSDEHGHEVHSPPDNHASHSDSDEAGSP
	orf46ng-1	: : : : : : : : : : : : NGHIGLGNIQSHOLGHIMIQQAAVEGNIGYIVRFSDEHGHKFHSPFDNHASHSDSDEAGSP
10		70 80 90 100 110 120
	orf46a.pep	130 140 150 160 170 180
	orf46ng-1	VDGFSLYIHWGDGYBHHPADGYDGPQQGGYPAPKGARDIYSYDIKVAQNIRNLTONRS
15		130 140 150 160 170 180
	orf46a.pep	TGQRQLDRFHNTGSMLTQVGQDGFKRATRYSPELDRSQNAAEAFNGTADIVRNIIGAAGE
	orf46ng-1	: : : : : : : : : : : TGQRLLADRFHNAGAMLTQVGQDGFKRATRYSPELDRSQNAABAFNGTADIVRNIIGAAGE
20		190 200 210 220 230 240
	orf46a.pep	IVGAGDAVQGISGNSNIAVNHHGLLSTENKMARINDLADMQLKDYAAAIRDWAVQNP
	orf46ng-1	: : : : : : : : : : IVGAGDAVQGISGNSNIAVNHHGLLSTENKMARINDLADMQLKDYAAAIRDWAVQNP
25		250 260 270 280 290 300
	orf46a.pep	NAAQGIEAVSNIFTAVIPVKGIGAVRKYGLGGITAHPVKRSQMGETALPKGKSAVDNF
	orf46ng-1	: : : : : : : : : NAAQGIEAVSNIFMAIPIKIGIGAVRKYGLGGITAHPVKRSQMGATALPKGKSAVDNF
30		310 320 330 340 350 360
	orf46a.pep	370 380 390 400 410 420
	orf46ng-1	ADAAYAKYPSPYHSRNIIRSNEQRYGEKNTSTSPTVPPSNGKNVKLJNRHPKTKVPFDGK
35		370 380 390 400 410 420
	orf46a.pep	ADAAYAKYPSPYHSRNIIRSNEQRYGEKNTSTSPTVPPSNGKNVKLADQRHPKTKVPFDGK
	orf46ng-1	430 440 450 460 470
40		GFPNFEKVKHYDTINTAVPQVN---PIDEVFN--PKGVSGLHSWSITARIQYAKLP
	orf46ng-1	: : : : : : : : GFPNFEKVKHYDTKLD--IQLSGGGIPKAKPVFDAKPRMVEDRKLN-KLTTRQEVKNV
45		430 440 450 460 470
	orf46a.pep	480 490 500 510 520 530
	orf46ng-1	RQGRIRYIIPPKNYPSAPlPKPGNNGYLDKEGNEWTKGPSRTKGQEFEDVQLSKTGREQ
50		: : QETRRSSQSQQFKAHAQREWEENRITGLDFNFIIGGDIINKGTVTGGHS1TRGDVVRVIQQT
	orf46a.pep	480 490 500 510 520 530
	orf46ng-1	801 CACCGAAAC AAGATGGCG GCATCACAGC TTTGGCAGAT ATGGCGCAC

The complete length ORF46a DNA sequence <SEQ ID 465> is:

55	1	TTGGGCATT	CCCGAAAT	ATCCCCTATT	CTGTCATAC	TGGCAGTGT
	51	CCTGGCGATG	CATGCCAACG	CTTCAGATTG	GGCAACAGAT	TCTTTTATCC
	101	GCGAGTTCT	CGACCGCTAG	CATTGCGAAC	CGGACGGGA	ATACCACTTA
	151	TTGGCGAGCA	GGGGGGGAACCT	TGCGGAGGCGC	AGCGGTCTATA	TCGGATTGGG
	201	AAACATACAA	AGCCATCACT	TGGGGCACCC	GTTCATCCAG	CAGCGCGCCA
	251	TTAAAGGAAA	TATCGCTAC	ATTGTCGGCT	TTTCGGATCA	CGGGCACCGAA
	301	GTCCATTCCC	CTTGTGCAAA	CTATGCCCTA	CATTGCCATT	CTGATGAAGC
	351	CGGTAGTCCT	GTGAGCGAT	TCACGGCTTA	CGCGCATCCAT	TGGGACGGAT
	401	ACGACACCCA	TCCCGCGAC	GGCTATGAGC	GGCCACACGG	CGGGCGCTAT
	451	CCCGCTCCCA	AAGGGCGAG	GGATATATAC	AGCTACGACA	TAAAAGGGCT
	501	TGCCAATAAT	ATCCCTCTCA	ACCTGGCGCA	CAACCGCACC	ACCCGACAC
	551	GGCTTGTGGA	CGTGTTCAC	AATACCGTGA	GTATGCTGAC	CGAACGGAGTA
	601	GGCGACGGAT	TCAAACCGCG	CACCCGATAC	AGCCCCGAGC	TGGACAGATC
	651	GGGAATATGCG	GGCGAACGCTT	TCAACCGCAC	TGCAAGATAC	GTCAAAAACA
	701	TCATCGGGCG	GGCGACGAGAA	ATTGTCGGG	CAGGGGATGC	CGTGCAGGGT
	751	ATAAGGGGAA	GCTCAACAT	TGCTGTATG	CACGGCTTGG	TGCTGTTTC
70	801	CACCGAAAC	AAGATGGCG	GCATCACAGC	TTGGCAGAT	ATGGCGCAC

851 TCAAAGACTA TGCGCGACCA GCCATCCGGG ATTGGGCAGT CAAAAACCCC
 901 AATGGCGCAC AAGGCATAGA AGCCGTCAGC AATACTTTTA CGGCAGTCAT
 951 CCCCGTCAAA GGGATTGGAG CTGTCGGGG AAAATCAGGC TTGGGGGCA
 1001 TCACCGCACCA TCTGTCAAG CGGTCCAGA TGGGGCAGAT CGCATGGCG
 1051 AAAGGGAAAT CGCGCTCA CGACATTTT GCAGATCGG CATAAGCCAA
 1101 ATACCCGTCG CCTTACCAATT CGCGAAATAT CGGTTCACAA AC TTGGAGCAGC
 1151 GTTACCGCAAG AGAAAACATC ACCTCTTCAG CGGTGCGGC GTCAAACCGA
 1201 AAGATGTGA AACITGGCAAA CAAACGCCAC CGGAAGACCA AAGTGCCTT
 1251 TGACGGTAAAG GGTTTCCGA ATTITGAAAC AGACGTTAAAG TAGCATACGA
 1301 GAATTAATAC CGCTGTACCA CAAGTGAATC CTATAGATGA ACCCGTCCTT
 1351 AATCTTAAGG GTTCTGTCC ATCGCTCAT TCTTGGCTCA TAACGTCCAG
 1401 AATTCAATAC CGCAAAATAC CAAGGCAAGG TAGAACTCAGA TATATCACC
 1451 CTAAAATTA CTCTCTTCAG CGACCGCTAC CAAAAGGACC TAATAATGGA
 1501 TATTGATTA ATTGTGTA TGAATGGACT AAAGGTCCAT CAAGAACTAA
 1551 AGCTCAAGA TTGGAATGGG ATGTTCAATC GTCTAAACAA GGAAGAGGAC
 1601 AACTTGGATG GGCTAGTAGG GATGGTAAAC ATTTAATAT ATCAATTGAT
 1651 GGAAGTTA CACACAAATG A

This corresponds to the amino acid sequence <SEQ ID 466>:

20 1 LIGISKISL LILAVCLPM HAHASLDLAND SFIRQVLDRQ HFPFDKYHL
 51 FGSREGLAEER SGHIGLNIQ SHOLGNLFID QIAIKNIGY IVRFSDHGH
 101 VHSPFDNHAS HSDESDENGSP VDFGSLYRIN WDGYEHHHIPD GYDGPQGGGY
 151 PAPKGARDIV SYDKIVGAQNL ARLNLTDRS TQQLRVLDRFH NTGSLMTQGV
 201 GDGFKRATRY SELDRSGNA AEAENGTADVI VKNIIIGAAGE IVVGADDAVQG
 251 ISEGENIAIVM HGCLLSTEN KMARINDLAD MAQLKDYYAA AIRWAVONP
 301 NAAQGEAIEVS NIIFTAVIPVK GIGAVGRKYG LGGITAHVPK FSGMGEIALP
 351 KGKGSVASDNF ADDAYAKYPS FYHSRNIRSN LEORYGENI TSSTVPSNG
 401 KNVKLANKRRH PTKTKVPTDGK GFENFEKDVK YDTRINTAVE QVNPIDEPVF
 451 NPKGVSVAHS SWSITARIQY AKLPRQRGRIR YIPPKNYSPS APLPKGPNNG
 501 YLDKGNEWNT KGPFRTHQGE FEWDVQLSKT GREQLGWASR DGKHLNISID
 551 GKITHK*

Based on this analysis, including the presence of a RGD sequence in the gonococcal protein, typical of adhesins, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

35 Example 56

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 467>:

40 1 ATGAATATTG ACACCCCTGCT CTCCAAACAA TGGACGCTGC CGCCATTCT
 51 CGCGAACCGG CTGCTGCTGT CCGTCTGAT ACTGCTTGCC CCGAATGGGG
 101 TGTTTGGGT TTGGCCTACT CGTACAGGCCA CGGCCGCCCG GATTGTCAT
 151 TTGGACTATC TTCCCGCCG GCTGCTGATC GCGCTTCCTT GGGCTTTCGT
 201 CAAATATGCC CGCTGATTTGG CGTTTGGCT GCGCTTGGT TTTGACGGGC
 251 TGATGATGGT GATCCAACTC TTCCCTTTA TGGATCTCAT CGGGCCATC
 301 AACCTCGTCC CCTTCATCCT GACCGCCCCC GCGCTTATC AGATAATGAC
 351 CGGGCTG...

45 This corresponds to the amino acid sequence <SEQ ID 468; ORF48>:

1 MNIIHTLLSKQ WTLPFLPLKR LLISLILILLA PNAVFVWLAL LTATARPIVN
 51 LDLYPAALLI ALPWFVFKIA GVLAFLWALV FDGLNMVIQL FPFMDLIGAI
 101 NLVPFILTAP APYQIMTGL...

Further work revealed the complete nucleotide sequence <SEQ ID 469>:

50 1 ATGAATATTG ACACCCCTGCT CTCCAAACAA TGGACGCTGC CGCCATTCT
 51 CGCGAACCGG CTGCTGCTGT CCGTCTGAT ACTGCTTGCC CCGAATGGGG
 101 TGTTTGGGT TTGGCCTACT CGTACAGGCCA CGGCCGCCCG GATTGTCAT
 151 TTGGACTATC TTCCCGCCG GCTGCTGATC GCGCTTCCTT GGGCTTTCGT
 201 CAAATATGCC CGCTGATTTGG CGTTTGGCT GCGCTTGGT TTTGACGGGC
 251 TGATGATGGT GATCCAACTC TTCCCTTTA TGGATCTCAT CGGGCCATC
 301 AACCTCGTCC CCTTCATCCT GACCGCCCCC GCGCTTATC AGATAATGAC
 351 CGGGCTGTTG CTGCTGATA TGGCTGGCAT GCGCTTGTG TTGAGAAAAG

401 CGCGCGCCAA CAGGACATTC CGGCACATTG CGGTCTGCC CGCCGTTG
 451 CGCGCAGCGC GCTATTTCAC CGGGCATTTTG AGTTTACTAG ACCGGGGTGC
 501 GATGGCAAT ATCTCGGCG AACAAACATT CTACTACGCC AAAAGTCAGG
 551 CGATGCTCA CACCGCTGAG CAAATGCGC ACTTTTATTAC CGCGCGCTG
 5 601 GTCGATCCCCTCCTGGCAAT CAACAGCGTGC CGCCACCGCA
 651 TCTGAACGAG CGGAATCTC AAAAATCTCTT TTATATGTC CGCGAATCTT
 701 GGGGGCTGCC GGCAAACTTC GAACCTTCAA ACGCCACTT TGCCAAACTG
 751 CTGGGGCAAA AAGGCCGTTT CGGAAATGCC CGAACCTGTG GCTTACGGCG
 10 801 CATCGGGCGC AGCGTGCAG CGGAAATGCC CGAACCTGTG GCTTACGGCG
 851 GTTTCGGCGG GTTGCACACT CGGCGCGGC CGGACGAAAA ATTGCCCCGC
 901 TGCCCTCCC ACCTTTTGA ACAGAAGGT TACGCCACTT TTGCGATGCA
 951 CGGGCGGGGC AGTTGCGTTT ACGACCGCTT CAGCTGGTAT CGGAGGGGGG
 1001 GCTTTCAAGA ATACAAAAGG GCGGAAACCC GCGATCGGTA AAAACCTGC
 1051 GCCATTTCGCGGCGCGAG CGACACGGAG CTGTTGCGGC AAGTGTGGC
 1101 ATTTTCAAA AAACAGCACAA AGGGACTTT TTACTTGATG ACCTGACCA
 1151 GCCACCGCGA CTATCGCGA TCCGACATT TCAACACAGG GCTCAAATGC
 1201 ACCGAATATG GCGTGGCGGC CGAACCCGAC CTCTGCCGCA ATTTCAGCT
 1251 GCACACCAAA TTCTCGAC AACTGGCGA TTGATGCAAA CGCCCGAAA
 1301 TGAAAGGCAC GGAAGCATC ATGTCGGCC ACCATCGGCC GCGCGTGGC
 1351 AACCTCAATG AACCTCCCG CTACCTCAAA CAGGGCGAGC TCACCTGGCT
 20 1401 GAACCTCAAA ATCAAAATAA

This corresponds to the amino acid sequence <SEQ ID 470; ORF48-1>:

1 MNIIHTLSKQ WTLPPFLPKR LLLSLLILLA PNAVFWVLAL LTATARPIVN
 25 10 LDLYPAALLI ALPWRFKVIA GVLAFLWALV FDGLMMVIOL FPFMMDLIGA
 101 NLVPLITAP APYOILMTGGL LYMLAMPEV LQKAAAKATDF RHAIVCAAVV
 151 AAAGYETGHL SYVDRGRMAN IPGANNFYYA KSQLMILYTVS QNADFITAGL
 201 VDPVFLPLGN QORATHLNE PKSQKILFIV AESWGLPANP ELONATFAKL
 251 LAQKDRFSVW ESGSFPTPIGA TVEGEMRELC AYGLGRFAL RRAPEKEFAR
 30 301 CLPNRLKQEG YATFMAHHGAG SSLYDRFSWY PRAGFQEIKT AENLIGKKTC
 351 AIFGGVCDFSE LFGEVSAAFK KHDKGLFYWM TLTSHADYPE SD1FNHRLKC
 401 TEYGLPAETD LCRNFSLHTQ FFQLQADLIQ RFENKGTEVI IVGDHPPPVG
 451 NLNETFRYLQ QGHVAWLNFK IK*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N. meningitidis* (strain A)

35 ORF48 shows 94.1% identity over a 119aa overlap with an ORF (ORF48a) from strain A of *N. meningitidis*:

	10	20	30	40	50	60
orf48.pep	MNIHHTLSKQWTLPPFLPKRLLLSSLILLAPNAVFVVLALLTATARPIVNLDYLPAALLI					
orf48a	MNIHHTLSKQWTLPPFLPKRLLLSSLILLXPNAVFVVLALLTATARPIVNLDYLPAALLI	10	20	30	40	50
		70	80	90	100	110
orf48.pep	ALPWRFKVIAVLAFWALVFDGLMMVIOLFPFMMDLIGAIVLVPFLTAPAPYQIMTGL	10	20	30	40	50
orf48a	ALPWXRVKIXGVILAXWLAIVLFDTGMMVIOLFPFMMDLIGAIVLVPFLXTPAPALYQIMTGL	70	80	90	100	110
		120				
orf48a	LLYMLAMPFLQKAAAKTDFRHIAACAAVVAAGYFTGHLSSXYDRGRMANIFGANNFYYA	130	140	150	160	170
						180

The complete length ORF48a nucleotide sequence <SEQ ID 471> is:

1 ATGATATTTC ACAGCCCTGT CTCCAACAA TGGACGCTGC CGCCATTCT
 51 GCGAACACGG CTGCTGCTG CCGTCTGAT ACTGCTNNCC CCGAATGCGG
 101 TGTATTTGGGTT TTTCGCACTG CTGACGCCA CGCGCGCCCG GATTGTCAT
 151 TTGGATTACCCCTGGCGC CTGCTGATC GCGCTGCTT GGGGNTCTG
 201 CAAATGTCNC CGCTCATTTG CGTGTGTTGC CGCGCTTGTG TTGACGGGC
 251 TGTGATGTCG GATCCAACTC TTCCCTTTA TGAGATCTCAT CGCGCCATC
 301 AACCTGCTACCTCCTTCACNT CGCGCCCGCC GCGCTTTATC AGATAATGAC
 351 CGGGCTGTAA CTGCTGATA TGCTGGCAT CGCGCTTGTG TTGCAAGAAG
 401 CGCGCGCCAA AACGGACTTC CGACACATTG CGCGCTGTCG CGCGCTGTC
 451 GTGGCAGCGC GCTATTTCAC CGGGCATTTG AGTTANTACG CGCGGGGGCG

501 GATGCCAAT ATCTTCGGG CAAACAACTT CTATTAAGCC AAAAGTCAGG
 551 CGATGCTCA CACGTGAGC CAGAATGCG ACCTTTATTAC CGCCGGCCTG
 601 GTGATCCCG TCTTCCTCCT CTGGGCAT CAAACAGGTG CGCCACCGA
 651 TCTGAACGAG CGGAATCTC AAAAATCTC TTATATGTCG GCCGAATCTT
 5 701 GGGGTGTCG GGGCAATCCC GAACCTCAA AGCCCAACTT TGCCAAACTG
 751 CTGGCCAAA AAAGANCGTT TTGGTTGG GAAGACGCCA GTTTTCCCTT
 801 CATCGGCCA AGCATCGAAAG CGCAAAATCG CGAACCTGTT GCCTACGGCG
 851 GTTGGCGGG GTTGGCACTG CGCCGGCGG CGCACGAAAAA ATTGGCCCG
 901 TGCTCCCC ACCCTTTGAA ACAAGAAGT TAGCCACCT TTGCGATGCA
 951 CGGCCGGGG AGTTGGTTT ACGACCGCTT CAGCTGGTAT CGGAGGGGG
 1001 GCTTCAGA AATCAAACCC GCGCAAAACCG TGATCGTTAA AAAACCTGC
 1051 GCCATTTCG GCGGCCTGTC CGACAGCGAG CTGTTCCGGG AAGTGTCCGC
 1101 ANTTTCAA AAACACGACA AGGGACTTT TAATCTGGATG ACCCTGACCA
 1151 GCCACGCCG ATCTACCGAA TCNGACATT TCAACACAG GCTCAAATGC
 1201 ACCGATATG GCTTCGCCGC CGAACCGAC NTCTGCACGA ATTTCAGCCT
 1251 GCACACCAA CCTTCGACCC AACTGGCGGA TTGATCTCAA CGCCCGGAAA
 1301 TGAAGGAC GGAAGTCATC ATCGTCGGCG ACCATCCGCC GCGCGTCGGC
 1351 AACCTCAATG AACCTCTCG CTACCTCAA CAGGGCAGG TCGNCTGGCT
 1401 GAACCTCAAATC AATCAAATAA

20 This encodes a protein having amino acid sequence <SEQ ID 472>:

1 MNIHTLLSKQ WTLPPFLPKR LLLSLLILLY PNAVFWVLAL LTATARPIVN
 51 LYXLPFAALLT ALPWRXVKIX GVLAIXLAVL FIDGIMMVQL FPFMDLIGI
 101 NLVPIFXTAP ALYQIMTGLL VLGMLAMPEFV LOKAAAKTDF RHTAACAAVV
 151 VAAYGFTGHL SKYDRGMAN IFGANNNFYA KSQAMLYTVS QNADFITAGL
 201 VDPVFLPLGN QORAAATHLNE PKSQKILFETIV AESWGLPANP ELQNATFAKL
 251 LAQKCRFSVW ESGSFPIFG TIREGEMRELIC AYVGLGRFAL KRAPDEKFAR
 301 CLPNRLKGEG YATFAMHGAG SSSLYDRFWY PRAGFQEIKT AENLIGKKTC
 351 AIFGGVCDSE LFGEVSAXXFK KHDKGLFYWM TLTSHADYPE SDIFNHRLKC
 401 TEYGLPAETD XCRNFSLHTQ FFDQLADLIQ REFMKGTEVI IVDHPPVG
 451 NLNETTRYLK QGHVXNLNFK IK*

ORF48a and ORF48-1 show 96.8% identity in 472 aa overlap:

	10	20	30	40	50	60
orf48a.pep	MNIHTLLSKQ WTLPPFLPKR LLLSLLILLY PNAVFWVLAL LTATARPIVN					
35 orf48-1	MNIHTLLSKQ WTLPPFLPKR LLLSLLILLY PNAVFWVLAL LTATARPIVN					
	10	20	30	40	50	60
	70	80	90	100	110	120
40 orf48a.pep	ALPWRXVKIXGVLAIXLAVL FDGLIMMVQL FPFMDLIGA INLVFFITAPALYQIMTGLL					
orf48-1	ALPWRXVKIAVGLAVL FDGLIMMVQL FPFMDLIGA INLVFFITAPALYQIMTGLL					
	70	80	90	100	110	120
	130	140	150	160	170	180
45 orf48a.pep	LYYMLAMPFLQKAAAKTDFRHIAACAAVVAAGYFTGHLISXYDRGRMANIFGANNNFYA					
orf48-1	LYYMLAMPFLQKAAAKTDFRHIAACAAVVAAGYFTGHLISXYDRGRMANIFGANNNFYA					
	130	140	150	160	170	180
	190	200	210	220	230	240
50 orf48a.pep	KSQAMLYTVSQNADFITAGLVDPVFLPLGNQQRAATHLNEPKSQKILFIVAESWGLPANP					
orf48-1	KSQAMLYTVSQNADFITAGLVDPVFLPLGNQQRAATHLNEPKSQKILFIVAESWGLPANP					
	190	200	210	220	230	240
	250	260	270	280	290	300
55 orf48a.pep	ELQNATFAKL LAQKCRFSVWESGSFPPIGATIEGEMRELICAYGLRGFALRRAPEKFAR					
orf48-1	ELQNATFAKL LAQKCRFSVWESGSFPPIGATIEGEMRELICAYGLRGFALRRAPEKFAR					
	250	260	270	280	290	300
	310	320	330	340	350	360
60 orf48a.pep	CLPNRLKGEGYATFAMHGAGSSLYDRFSWYPRAGFQEIKTAENLIGKKTCAIFGGVCDSE					
orf48-1	CLPNRLKGEGYATFAMHGAGSSLYDRFSWYPRAGFQEIKTAENLIGKKTCAIFGGVCDSE					
	310	320	330	340	350	360

-280-

		370	380	390	400	410	420
	orf48a.pep	LFGEVSAXFKKHDKGLFYWNNTLTSADYPPESIDFNHRLKCTEYGLPAETDXCRNFSLHTQ					
5	orf48-1	LFGEVSAXFKKHDKGLFYWNNTLTSADYPPESIDFNHRLKCTEYGLPAETDXCRNFSLHTQ					
		370	380	390	400	410	420
		430	440	450	460	470	
10	orf48a.pep	FFDQLADLIRQPEMKGTEVIIVGDHPPVGNLNNETFRYLKQGHVXWLNFKIKX					
	orf48-1	FFDQLADLIRQPEMKGTEVIIVGDHPPVGNLNNETFRYLKQGHVXWLNFKIKX					
		430	440	450	460	470	

Homology with a predicted ORF from *N.gonorrhoeae*ORF48 shows 97.5% identity over a 119aa overlap with a predicted ORF (ORF48ng) from *N.*15 *gonorrhoeae*:

	orf48.pep	MNIHTLLSKQWTLPPLPKRLLLSSLILLAPNAVFVWLALLTATARPIVNLDYLPAAILLI	60
	orf48ng	MNIHALLSEQWTLPPFLPKRLLLSSLILLAPNAVFVWLALLTATARPIVNLDYLPAAILLI	60
20	orf48.pep	ALPWRFVKIAIGVLAFWLAFLFDGLMMVIQLFPFMDDLGAINLVPFILTAAPAPYQIMTGL	119
	orf48ng	ALPWRFVKIAIGVLAFWPAVLFDGLMMVIQLFPFMDDLGAINLVPFILTAAPAPYQIMTGL	120

The ORF48ng nucleotide sequence <SEQ ID 473> was predicted to encode a protein having amino acid sequence <SEQ ID 474>:

25	1	MNIHALLSEQ WTLPPFLPKR LLLLSSLILLAPNAVFVWLAL TATARPIVN	
	51	LDLYLPAAILL ALPWRFVKIA GVLAFWPAVL FDGLMMVIQLFPFMDDLGAINLVPFILTAAPAPYQIMTGL	
	101	NLVFEITAP APYQIMTGLL LYLMALMFVY LOKAAVRTDF RHIAVCAAVV	
	151	AAARYTGP ELLRTGGRWW YVQHRRLLLS GSRSASFRQQ KADVLERLGN	
	201	PYASMSNGG ..	

30 Further work identified the complete gonococcal DNA sequence <SEQ ID 475>:

	1	ATGAATATTG ACGCCCTGCT CTCGGACAAA TGGACGCTGC CGCCATTCCCT	
	51	CGCCAAACGG CTGCTGCTGT CCTCTGCTAT ACTGCTGGCC CCCAACGGCG	
	101	TGTTGCTGGT TTTCGCTCTG CTGACGCCCA CGCGCCGCCCG GATTTGCGAT	
	151	TTCGACTTACCC TTTCGGCCCG CGCTCTGATC CGCCCTGGCTT CGCGTTTCTT	
	201	CAAATTGCC CGCGTATGG CGTTTTGGC GCGGGTTTGG TTGACGGGC	
35	251	TGATGATGGT GATCTCTACCT TTCCCTTTTA TGACCTCAT CGCGCCGCATC	
	301	AACCTCGGCC CCTTCATCCT GACGCCCGCC CGCCCTTATC AGATAATGAC	
	351	CGGGCTGTGT CGCTGTGATA TGCTGGCGAT GCGCTTGTG TTGCAAAAG	
	401	CGCCGCTCAC AACCGACTCG CGACACATTG CGCTGTGCG CGCGCTGTG	
	451	CGCGCACGCG CCTATTTACCG CGGCCATTGTT AGTTTACTACG ACCCGGGGGCG	
	501	GATGGCCCAAG ATCTTCCCGCC CAAACAACTT CTATTACGGC AAAAGTCAGG	
	551	CGATGCTCTA CACCGTCAGC CAGAACGGCG ACCTTTATTA CGCGGCGccTG	
	601	TGCGACCCCG CCTTCCTCC TTTGGCCAGT CAGCAGGGTG CGCGCACCGG	
	651	GCTGATGAG CGGAAATCTC AAAAATTCCTT CTTATATGCTG CGCGAAATCTG	
45	701	GGGGCGCTCC GGGCAATCTC GAGCTTCAAA CGCCCACTTT TGCCAAACTTG	
	751	CTGGCCCAAAG AAAGCCCTT TTGGCTTTGG GAAAGCGCGA GTTTTCCCTT	
	801	CATGGCCCGG CGGGCTCAAG CGCAATTCGG CGAAATTGTGC CGCTTACGGCG	
	851	GTGTTGGCGGG GTTGGCACTG CGGCCGGCGGC CGCGACGAAAA ATTTCGCCCC	
	901	TGCTTCCCCC ACCTGTTGAA ACAAGAAGGT TGCCTACCTT TTGGATGTC	
	951	CGGGCGGGT AGCTCGCTT ACCGACCGCTT CGCTGGTGTAT CGGAGGGGG	
50	1001	GCTTCAAAAT AATCAAAACG CGCGAAARACG TGATGCTTAA AAAAACCTGC	
	1051	GGCCATTTCG CGGGCGCTGT CGACAGCGAG CTGTTCCGGCG AGTGTGTC	
	1101	ATTTTCAAAAT AAACACGACA AGGGACTGT TTFACTGGATG ACGCTGACCA	
	1151	GGCACCGCGA CTATCCGAA TCCGACATTG TCAACACAG CGCTAAATGC	
	1201	ACCGAAATAGC GGCTGCCCGC GAACACCGAC TCTCTGGCGA ATTTCAGCGCT	
	1251	GCACACCCAG Tttcttcggacc AACTGGCGGC TTGATTCGA CGCCCCGAAA	
	1301	TGAAGGCAGC GGAAGTCATC ATGCTGGCG ACCATCCGGC GCGCGTCGGC	
	1351	AACCTCAATG AAACCTTCGG CTACCTCAA CAGGGACACG TCGCCCTGGCT	
55	1401	GCACCTCAAATCATAATAA	

60 This encodes a protein having amino acid sequence <SEQ ID 476; ORF48ng-1>:

1	MNIHALLSEQ	WTLPFFPLPKR	LLLSLLILLIA	PNAVFWVLAL	LTATARPIVN
51	LDYLPAAALLI	APEWRFVKIA	GVLAFWAVL	FDGLMMVQL	FFFMIDLIGAI
101	NLVFPIITAP	AQPYQIMTGL	LILYMLAMPF	LQRAAVKTF	RHIAVCAAVV
151	AAAGYFTGHL	SYYDRGRMAI	IRGANNNFYA	KSQQMLTYVS	QNADFITAGL
201	VDPVFPLGN	CQRARTLSE	PSKSQKILEIV	AESWLGPNP	ELQNATPAKL
251	LAQKDPRFSWV	EGCSFFFGIA	TVEGEMRELIC	AYGGRLGFAL	RRADFEKFAR
301	CLPNRLKQEC	YATFAMHGAG	SSLYDGRFSWY	PRAGFQKIKT	AENLIGKKT
351	AIFGGVCDSE	LFGEVSAAFP	KHDKGFLFYWM	TLTSHADYPF	SDIFNHLKC
401	TEYGLPAETD	LCRNFSLHTQ	FFDQLADLR	RPEMKGTEVI	IVGDHPPVG
451	10	NLNETFRYLK	QGHVAWLHFK	IK*	

ORG48ng-1 and ORF48-1 show 97.9% identity in 472 aa overlap:

	10	20	30	40	50	60
orf48-1.pep	MNIHALLSKQWTLPPFLPKRLLLSSLILLAPNAVFWVLALLTATARPIVNLDYLPALLI					
15 orf48ng-1	MNIHALLSEQWTLPPFLPKRLLLSSLILLAPNAVFWVLALLTATARPIVNLDYLPALLI	10	20	30	40	50
						60
orf48-1.pep		70	80	90	100	110
20 orf48ng-1	ALPWRFVKIAVGVLAFWIAVLFDFGLMMVIQLEPFPMDLGIAINLVPFILITAPAPYQIMTGL	10	20	30	40	50
						60
orf48-1.pep		70	80	90	100	110
25 orf48ng-1	ALPWRFVKIAVGVLAFWIAVLFDFGLMMVIQLEPFPMDLGIAINLVPFILITAPAPYQIMTGL	70	80	90	100	110
						120
orf48-1.pep		130	140	150	160	170
30 orf48ng-1	LLYMLAMPFLQKAAAKTDFRHIAVCAAVVAAGYFTGHLISYYDRGRMANIFGANNFYA	130	140	150	160	170
						180
orf48-1.pep		130	140	150	160	170
35 orf48ng-1	LLYMLAMPFLQKAAAVKTDDFRHIAVCAAVVAAGYFTGHLISYYDRGRMANIFGANNFYA	130	140	150	160	170
						180
orf48-1.pep		190	200	210	220	230
40 orf48ng-1	KSQAMLYTVSQNADFITAGLVDPVFLPLGNQQRATHLNPEKSQKILFIVAESWGLPANP	190	200	210	220	230
						240
orf48-1.pep		190	200	210	220	230
45 orf48ng-1	KSQAMLYTVSQNADFITAGLVDPVFLPLGNQQRATHLSEFKSQKILFIVAESWGLPNP	250	260	270	280	290
						300
orf48-1.pep		250	260	270	280	290
50 orf48ng-1	ELONATFAKLLAQDKDRFSWESGSFFFIGATVEGEMRELICAYGGLRGFALRRAPEKPAR	250	260	270	280	290
						300
orf48-1.pep		310	320	330	340	350
55 orf48ng-1	ELONATFAKLLAQDKDRFSWESGSFFFIGATVEGEMRELICAYGGLRGFALRRAPEKPAR	310	320	330	340	350
						360
orf48-1.pep		310	320	330	340	350
60 orf48ng-1	GLPNRLKQEGYATFAMHGCGSSLYDRFSWYPRACFQEIKTAENLICKKTCGAFIVGGVCDSE	310	320	330	340	350
						360
orf48-1.pep		310	320	330	340	350
65 orf48ng-1	CLPNRLKQEGYATFAMHGCGSSLYDRFSWYPRACFQEIKTAENLICKKTCGAFIVGGVCDSE	370	380	390	400	410
						420
orf48-1.pep		370	380	390	400	410
70 orf48ng-1	LPGEVSAAFFKKHDKGLFYWMILTSHADYFESDIFTNHLRKCTEYGLPAETDLCRNFSLHTQ	370	380	390	400	410
						420
orf48-1.pep		430	440	450	460	470
75 orf48ng-1	FFDQLADLIRPEMKGTEVIIVGDHPPFVGNLNETFRYLIKQGHVAWLNFKIK	430	440	450	460	470

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and two putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 57

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 477>:

```

      1 .. GTGAGCCGGAC GTTACCGGGC TTGGATCGC GTTCCAAAAT TCATCATGCT
  5   51 TACTTGAGT ATCCCCACCG TTGGCGGCC CGGCATCGT ATGCGGCC
1001 101 GTATGCGAGAT GCAGCTCCGAT TTATCGAGC CGCACACCGTG GACGCTTGC
151 151 GGTTGGGCT TCCGTGTCG GCTGATGGC TGATGCC CGCCGATTGA
201 201 AATTTCCGCGC ATCAATTCTT TGTTGGTAAC CGAAARACAA CGCATCAATC
251 251 CTTCGGAAATA CGCGACGG AGTTGAAT TCANCGTGG TTATATCGCC
301 301 AGTGGGGTTT TGCGTTGTT TTTCCTTGTCA CTGGGGC G TAGCCCGAA
351 351 CGCGAACGGC GA ACAGTGC AGATGGCGG CGGAAATAT AACGGGCAAT
401 401 TGATCAATAT GTACGCC..

```

This corresponds to the amino acid sequence <SEQ ID 478; ORF53>:

```

      1 .. VSGRYRALDR VSIIIVTLS IATLAAAGIA MSRGMMQMSDF FIEPTFWTLA
  5   51 GLGFLIALMG WMPAIEISA INSLWTEKQ RINPSEYRDG IFEFNVGYIA
1001 101 SAVLALVFLA LGXVAEPNGN XT'VQAGGKY NGQLINMYA...

```

Further work revealed the complete nucleotide sequence <SEQ ID 479>:

```

      1 ATTCGGAAAC AACATATTTT GACTTGGAAA AGTAAATATCA ACGCATGGG
  5   51 TCCGGGATC ATGATGCGT CGCGGGGGT CGCGGTTTG CACCTGATG
1001 101 ATCCGAGGCA GGCGGGCCCG CTTACCGGT GGCGATGTCG GCTCATCATC
151 151 ATCCGTGACCA ACCCTCTTCA ATACCCTTT TTCCGCTTCA CGGGCACTTA
201 201 CACCTGCGAC AGCGGCAAGA GCCTGATGGA AGGTTATGCC GAGAAAAGCC
251 251 GCGTTTATTATC TTGGGTATTC CTGATTTTG GCACTCTC CCCACGATT
301 301 AACGGGGGGG GGTCGSCCAT TGTAACCGCC GGCATGTCG AAATGGCGAT
351 351 TCCCTGGCTG ATGTTTGATC CGGGCACGGT TGCGGCTTG ATTATGGCAT
401 401 CCTGGCTGAT TATTGGTGTG AGCCGGACGGT ACCGGGCTTG GGATGGGTT
451 451 TCCAAATATCA TCATGGTAC TTTGAGTATC GCAACGCTG CGCGGGCGG
501 501 CATGGCTGATC TCGGGGGGTG TGCAAGTGC GTCCGATTTT ATCGAGGCCG
551 551 CACGGTGGAC GCTGGCGGT GCTGGCTTC TGATGGCCT GATGGGCTGG
601 601 ATGGCCGGC CGATGAAAT TTCCGGCCATC AATTCTTGT GGTTAACCGA
651 651 AAAAACACGC ATCAATCTT CGCAATACCG CGACGGGAT TTGATTTCA
701 701 ACCTCGGTTA TAFCGCCAGT GCGGGTTTGG CTTGGTTTT CTTGCACTG
751 751 GCGGGTTTG TGCAATACGG CAACGGCGAA GCAGTGCAGA TGCGGGCGG
801 801 CAATAATATC GGCACATG TGAAATATGA CGCGTGTAC ATCGGGCGCT
851 851 GGTGGCGCCC GTCTGGCGG TTATGGCGT TTGCGTGTAT FGTACGGCACG
901 901 AGCATATCCC TCGTGGACGG CTAATGGCGT GCAATGGCG AACCGTGGC
951 951 CCTGCTGCC CGGAAAGACA AAACGGGCAA CGCGGAATTC TTGCGTGG
1001 1001 ATATTGGGT GCGGGCGACG GGTGGCGG TGATTTCTG TTGTTAGGGC
1051 1051 GTATGGCA ATCTGGCAA ATTTGGCATG ATTGGCGCTT TTGTTGTCGC
1101 1101 CCCTGTGTT GCGCTGGCTGA ATTACCGTTT GGTTAAAGGT GTGAAAACAC
1151 1151 ACAAACTCAC ATACGGTATG ATGGCCCTG CATTGGCAGG TTGATTTAT
1201 1201 CTGACCGGGT TTACCGTTT GTTCTTATTG AATTGGCGG GATGTTCAA
1251 1251 ATGA

```

This corresponds to the amino acid sequence <SEQ ID 480; ORF53-1>:

```

      1 MSEQHISTWK SKINALPGI MMASAAGVGS HLIASTORGA LYWGQIALII
  5   51 ILTNLFKYPF FRFSAHYTLG TGKSLIEGTA EKSRVYLWVF LILCILSATA
1001 101 NAGAAVAVITA AIVKMAIFSL MFDAGTVAAL IMASCLLILV SGRYRALDRV
151 151 SKIIVTLSI ATLAAGIAA SRGMQMSDF IEPTFWTLAG LGFLIALMGW
201 201 MPFPIEISAI NSLWTEKQIN PNSEYRDGF DFENVGYIAS AVLALVFLA
251 251 GAFVQYGNGE AVQMAGGKYI QOLINMYAVT IGGWSRPLVA FIAFACMYGT
301 301 TITVWDGYAR AIAEPVRLLA GRDKTGNAEF FAWNIWVAGS GLAVIEWFWDG
351 351 VMANLKFAM IAAFVSAPF AWLYNRLVGK DEKHKLTSQM NALALAGLYI
401 401 LTGGTFLFLL NLAGMFK*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

- 55 ORF53 shows 93.5% identity over a 139aa overlap with an ORF (ORF53a) from strain A of *N. meningitidis*:

			10	20	30
	orf53.pep		VSGRYRALDRVSKIIIVTLSIATLAAAGIA		
5	orf53a	AAIVKMAIPSMLFDAGTVAAIMASCLILVSGRYRALDRVSKIIIVTLSIATLAAAGIA	T T T T T T T T T T T T T		
		110 120 130 140 150 160			
10	orf53.pep	MSRGQMOSDFIEPTPWTLAGLGFLIALMGWMPAPIEISAINSLLWVTEKQRINPSEYRDG	T T T T T T T T T T T T T		
	orf53a	MSRGQMOSDFIEPTPWTLAGLGFLIALMGWMPAPIEISAINSLLWVTEKQRINPSEYRDG	T T T T T T T T T T T T T		
		170 180 190 200 210 220			
15	orf53.pep	IPEFNVGYIASAVLALVFLALGXVAPNGNGXTVQAGGKYNGQLINMYA	T T T T T T T T T T T T T		
	orf53a	IPEFNVGYIASAVLALVFLALGAFVQYNGEAVQAGGKYIGQLINMYAVTIGGWSRPLV	T T T T T T T T T T T T T		
		230 240 250 260 270 280			
20	orf53a	AFIAFACMHTTTVVVDGYARAAEFPVRLRGDKTGNAEFFAWNIWVGSLAVIFWFD	T T T T T T T T T T T T T		
		290 300 310 320 330 340			

The complete length ORF53a nucleotide sequence <SEQ ID 481> is:

1	ATGTCGGAAC	AACATATTC	GACTTGGAAA	AGTAAATCA	ACGCATTGGG
51	ACCggggatt	ATGATGGCTT	CGGGGGCGGT	CGCCGGTTTC	CACCTGATTG
101	CCTGGACGCA	GGGGGGGGCG	CTTTACGGGT	GCGAGATGCC	GCTCATCATC
151	ATTCCTGACCA	ACCTCTTCAA	ATACCGGTT	TTCCGCTTCA	GGGGGCGATTA
201	CACCGTGAC	ACCGGCAAGA	GCTCTGATTA	AGGTTATGCC	GAGAAAAGCC
251	GGCTTATTAT	GTGGGTATTC	CTGATTITGT	CGATCCTCTC	CGCCAGGATT
301	AACCCGGGG	CGTCGCGCAT	TGTAACCGCC	GCCATCGTC	AAATGGCGAT
351	TCCCTGGCTG	ATGTTGATG	CGGCCACGGT	TGCCCCCTTG	ATTATGGCAT
401	CCTGGCTGAT	TATTGGTGT	AGCGGACGTT	ACCGGGCTTT	GGATCCGCTT
451	TCCAANATCA	TCACTGTTAC	TTGACTATAC	GCCACGCTTG	CGCGCCCGG
501	CATGGCTATG	TGGCGGGT	TGCAAGATGCA	GTGCGATTTT	ATCGAACCCGA
551	CACCGTGAG	GCTTCGCGGT	TTGGGCTTC	TGATGGCGCT	GATGGCTGG
601	ATGCGGGCG	CGATGGAAT	TTCCGGCCTA	AATTTCTTGT	GGGTAAACCGA
651	AAAACAAACGG	ATCAATCCTT	CGGAATACCC	CGACGGGATT	TTTGTATTCA
701	ACGTCTGGTA	TATCCGCACT	GGCTTTTGTG	CTTGTGTTTT	CTTGTGACTTG
751	GGCCGTTTTA	TGCAATCAGG	CAACGGCGGA	GCAGTGCAGA	TGGGGGCGG
801	CAATAATATC	GGCGCAATTGA	TCAATATGTA	GGCGCTTACCC	ATCGGGGGCT
851	GGTGGGGCCC	GCTTGGCGGT	TTATCGCGT	TTGCGCTGAT	GTAGGGCACG
901	ACGATTAACCC	TTGTGGACGG	CTATGCGCCGT	GGCAATTCGG	AAACCGTGGG
951	CCTGGCTGGC	GGAAAAGACA	AAACGGCGCA	CGCCGAATT	TTTGCCTGGA
1001	ATATGGGGT	GGGGGGCAGC	GGTTTGGCGG	TGATTTTCTG	GGTTGACGCC
1051	GTATGGCGA	ATATGCTCAA	ATTTGGCATG	ATGGCGCTT	TTGTGTCGCC
1101	CCCTGTGTT	GGCTGTGTA	ATTACCGTTT	GGTCAAAGGT	GATGAAAAC
1151	ACAATACAC	ATCAAGTATC	ATAGCCCTTG	CATTGGCAGG	CTTGATTTAT
1201	CTGACGGGT	TTACCGTTTT	GTTCCTTATG	AATTGGCGG	GAATGTCRA
1251	ATGA				

This encodes a protein having amino acid sequence <SEQ ID 482>:

1	MSEHQHISTWK	SKINALGPGL	MMASAAGVGS	HЛИASTQAGA	LYGWQIALII
51	ILTNLFKYFP	FRFSAHYTLD	TGKSLIEGYA	EKSRVYLWVF	LILCILSATI
101	NAGAVAIVTA	AIVKMAIPSL	MFDAGTVAAI	IMASCLILIV	SGRYRALDRV
151	SKIIIVTLSI	ATLAACAGIA	SRGMQMSDF	IEPTFWTLA	LGFILIALMW
201	MPAFIELSAI	NSLWTEKQR	INPSEYRDG	FDFNVGYIAS	AVLALVFLAL
251	GAFVYOGNGE	AVCMAGGKYI	GOLINNHYAVT	IGGWSPRELVA	FIACFAMYGT
301	TITWDGYAR	AAIEFVRLR	GKDRTGNAEF	FAWNIWVAGS	GLAVIEWFDG
351	VMANLFLKFM	IAAFVSAPCA	AWLNLYRVLKG	DEKHKLTSQM	NALALAGLIY
401	LTGFITVLFL	NLAGMF*			

ORF 53a shows 100.0% identity in 417 aa overlap with ORF53-1:

		10	20	30	40	50	60
60	orf53a.pep	MSEHQHISTWKSKINALGPGLIMMASAAGVGS	HЛИASTQAGA	LYGWQIALII	LTNLFKYFP		
	orf53-1	MSEHQHISTWKSKINALGPGLIMMASAAGVGS	HЛИASTQAGA	LYGWQIALII	LTNLFKYFP		
		10	20	30	40	50	60
65		70	80	90	100	110	120

-284-

	orf53a.pep	FRFSAHYTLDTGKSLIEGYAEKSRVYLWFLILCILSATINAGAVAIIVTAIAVKMAIPSL 	130	140	150	160	170	180
5	orf53-1	FRFSAHYTLDTGKSLIEGYAEKSRVYLWFLILCILSATINAGAVAIIVTAIAVKMAIPSL 70	80	90	100	110	120	
	orf53a.pep	MFDAGTVAAALIMASCLLIVSGRYRALDRVSKIIIVTLSIATLAAAGIAMSRRGMQMSDF 	130	140	150	160	170	180
10	orf53-1	MFDAGTVAAALIMASCLLIVSGRYRALDRVSKIIIVTLSIATLAAAGIAMSRRGMQMSDF 130	140	150	160	170	180	
	orf53a.pep	IEPTPWTLAGLGFLIALMGWMWPAPIEISAINSLSWVTEKQRINPSEYRDGIFDFNVGYIAS 	190	200	210	220	230	240
15	orf53-1	IEPTPWTLAGLGFLIALMGWMWPAPIEISAINSLSWVTEKQRINPSEYRDGIFDFNVGYIAS 190	200	210	220	230	240	
	orf53a.pep	AVLALVFLALAGFQVQYGNGEAVQMAGGRYIGQLINMYAVTIGGWGRPLVAFIAFACMYGT 	250	260	270	280	290	300
20	orf53-1	AVLALVFLALAGFQVQYGNGEAVQMAGGRYIGQLINMYAVTIGGWGRPLVAFIAFACMYGT 250	260	270	280	290	300	
	orf53a.pep	TITVVDGYARAIAREPVRLRKGDKDTGNAEFFAANWIWAGGGGLAVIEWFDGVMANLLKEAM 	310	320	330	340	350	360
25	orf53-1	TITVVDGYARAIAREPVRLRKGDKDTGNAEFFAANWIWAGGGGLAVIEWFDGVMANLLKEAM 310	320	330	340	350	360	
	orf53a.pep	IAAFVSAPVFAWLNRYRLVKGDDEHKHLTGKMNALALAGLIYLTFGTVLFLNLNLAGMFKX 	370	380	390	400	410	
30	orf53-1	IAAFVSAPVFAWLNRYRLVKGDDEHKHLTGKMNALALAGLIYLTFGTVLFLNLNLAGMFKX 370	380	390	400	410		
35	<u>Homology with a predicted ORF from <i>N. gonorrhoeae</i></u>							
	orf53a.pep	VSGRYRALDRVSKIIIVTLSIATLAAAGIA	30					
40	orf53ng	AAIVKMAIPSMLMFDAGTVAAALIMASCLLIVSGRYRALDRVSKIIIVTLSIATLAAAGIA	91					
	orf53a.pep	MSRGMMQMSDFIEPTPWTLAGLGFLIALMGWMWPAPIEISAINSLSWVTEKQRINPSEYRDG 	90					
45	orf53ng	MSRGMMQMSDFIEPTPWTLAGLGFLIALMGWMWPAPIEISAINSLSWVTEKQRINPSEYRDG	151					
	orf53a.pep	IEFENGVYIASAVLALVFLALGKVAPNGNGXTVQMGAGKYNGLINMYA	139					
50	orf53ng	IEFDENVGYIASAVLALVFLALGAFVQYGNGEAVQMGGGKYIGQLINMYAVTIGGGSRPLV	211					

An ORF53ng nucleotide sequence <SEQ ID 483> was predicted to encode a protein having amino acid sequence <SEQ ID 484>:

1	MPKKSCVYLM VFLILCIBSA TINAGAVAI TAAIVKMAIP SLIMFDAGTV
5	51 ALIMASCLLIVSGRYRALDRVSKIIIVTLSIATLAAAGI ANSRGMQMSDF
55	101 DFIPTPWTLAGLGFLIALMGWMWPAPIEISAINSLSWVTEKQRINPSEYRDG
	151 GIFDFNVGYIASAVLALVFLALGAFVQYGNGEAVQMGGGKYIGQLINMYA
	201 VTIGGGSRPL VAFIAFACMY GAATSTVVDGY ARAIAEPVRL LRKGDKTARP
	251 IVLEKLGLGR HRFGRDFL*

Further analysis revealed further partial DNA gonococcal sequence <SEQ ID 485>:

60	1 ...aaagaAAGAT GCGTTTTATT GTGGGGTTTT TTGATTTGT GTATCGCTC
	51 CGCACGATT AACGCCGGCG CGGTGCGCAT TGTAACCGCC GCCATCGTC
	101 AAATGGCGAT TCCCTCGCTG ATGTTGATG CCGGCACGGT TGCCGCTTG

-285-

151	ATTATGGCAT	CCIGGCCTGAT	TATTTTGGTG	AGGGGACCTT	ACCGGGCTTT
201	GGATCTGTG	TICCAAATCA	TCTATGGTAC	TTGAGGATC	CCACGGCTTG
251	CGCCGCCCGG	TACTCTGAT	TGCGGCGGTG	TTCAGATGCA	GGCGGAATTG
301	ATCGAGCCGA	CACCTGGAC	GCTTGGCGAT	TIGGCCCTTC	TGATCCGTTG
351	GATGGCTGAA	ATAGCCGGCC	GTATGGCAAT	TTCCTCCCGA	ATATTCTTGT
401	GGGTAAACCGG	AAAACACCGG	ATCATCTTC	CTGAATACCC	CGACGGGCTG
451	TTTCGATTC	ACGCTGGTTA	TATGCCGcat	GGCGTTTTCG	TCTGGTTT
501	CCTTGTGACTG	GGCCGGGTTG	TGCAATACCG	CAACGGCGAA	GAAGCTGACCG
551	TGGCGGGCGG	CAAATATATG	GGCAATATGA	TATATAATGTA	TCGGCTTACAGG
601	ATTCGGGGCT	GGTCTCTGTC	GTCTGGGCG	TITATCTGGCT	TTGGCTGTAT
651	GTACGGCACG	AGCATTCATG	TGGTGGACCG	TATGCGGCT	GTATGGCTCCG
701	AACCCCTGCG	CTTGCTGCCT	GGCAGGGATA	AAACCGGCAA	CGCCGAGTTG
751	TTTggCTGGA	ATATTGGGT	GGGGCGGACG	GGTGTGGCCG	TGATTTGCTTG
801	GTTTGAAGcg	gcaAGGCG	AACtgctTCA	ATTTCGGAG	TtttcgcgcT
851	TGTGCGGCG	CCCTGTTGTC	GGCTGGCTCA	ACTAACGGCT	CTGAAAGGG
901	GACAAACGGC	ACGACTTCA	GGGGCGGTAG	AAACGGCTTG	CAATTGTCG
951	CTCTGCTCTAC	TGCGGGGGT	TIGGCCCTTT	GTTCCTGTG	AACTTCTACCG
1001	GACTTGTGCG	ATAG			

This corresponds to the amino acid sequence <SEQ ID 486; ORF53ng-1>.

20 1 .KKS^CYLVWVF LILCIASATI NAGAVAI^VT AIVKMAIPS^L MFDGTAVAL
 51 IMASC^CLILWV SGRYALRDRV SKILLIT^VTSI ATLAACAGAM SGRGMQMPDF
 101 IEPTFWL^IAG FLGLIAWM MPAFIEISAI NSLW^TTEKQ^R INSYERDGI
 151 FDFN^VCY^G AVAFLV^CGF^AQVYNG^E AVQMGAKGY^I KNLINMYAVT
 201 IGGWSR^VLA VFIACFM^GYT TITV^DU^CY^GAR AIAE^FVRLR^R GRDKTN^GAL
 25 251 FAWN^CWI^GAS GLAVIFEW^DGA AMELA^KFKV^F IAAFS^AV^PE AWLN^YRLVK^G
 301 DKHR^LTAGM^M NALA^VGLV^GLAG FAVFL^L NL^TGLLA^R

ORF53ng-1 and ORF53-1 show 94.0% identity in 336 aa overlap:

		60	70	80	90	100	110
30	orf53-1.pep	ILTNLFKYFFFRSAHYTLTDGKSLIEGYAEKSRLVLYWFLILCILSATINAGAVAI					
	orf53ng-1			KKSCVYLLWFLILCIASATINAGAVAI			
35					10	20	30
	orf53-1.pep	120	130	140	150	160	170
35	orf53ng-1	AIVKMAIPLSMLFDAGTVAALIMASCLLIVSGRYRALARLDRVSRLKIIIVLTSIA					
40	orf53-1.pep	AIVKMAIPLSMLFDAGTVAALIMASCLLIVSGRYRALARLDRVSRLKIIIVLTSIA					
	orf53ng-1	40	50	60	70	80	90
45							
	orf53-1.pep	SRGMQMOSDFIEPTPWTLAGLGFIALMGMWMPAPIEISAINSLSWTEKQRINPSEYRDGI					
45	orf53ng-1	SRGMQMOSDFIEPTPWTLAGLGFIALMGMWMPAPIEISAINSLSWTEKQRINPSEYRDGI					
		100	110	120	130	140	150
50							
	orf53-1.pep	FDNVGYIYASAVLALVFLALGAFVQYGNGEAQVMAGKKYIGQLINMYAVTIGGSWSRLV					
50	orf53ng-1	FDNVGYIYASAVLALVFLALGAFVQYGNGEAQVMAGKKYIGQLINMYAVTIGGSWSRLV					
		160	170	180	190	200	210
55							
	orf53-1.pep	FIAFACMYGTTITVVVDGYARAIAEFVRLLRGKDGTGNAEFFAWNINIWAGSGGLAVIFWFDG					
55	orf53ng-1	FIAFACMYGTTITVVVDGYARAIAEFVRLLRGKDGTGNAEFFAWNINIWAGSGGLAVIFWFDG					
		220	230	240	250	260	270
60							
	orf53-1.pep	VMANLLKFAMIAAFVSAVPFAWLNYLRVKDEKKHLTSGMNALLAGLILYTGTFTVFL					
60	orf53ng-1						
		AMAEELLKFAMIAAFVSAVPFAWLNYLRVKDEKKHLTSGMNALLAGLILYTGTFTVFL					
60		280	290	300	310	320	330

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be

5 useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 58

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 487>:

```

1 .. TTGCGGGAAA CGGCATATGT TTTGGATAGT TTTGATCGTT ATTTGGTTGT
10 51 TGGCGTTGCC GGCTTGTGTT TTGTCGCCGC ACATCGGAA CGCAGCTGGA
101 101 TGGCGGAGGT TTCTCGCTGG CAGGAANAAA AACGGGAAAAA ACAGGGGGAC
151 151 CTGGCTGAAA TCAAAAGCGG TATGCCCGAT TTCCCGAC TTAGCTGTAT
201 201 GCTTTTCAC CGCGTCAAAGA CGCGAGTGTG TTGGCTGTGTT GTCCGGTGTG
251 251 CGCTTCTCTG CGCGAACATAT CGGGGGCGAC AATCCGAAACG GGACAGGCC
301 301 GTTCCGGCTT .

```

15 This corresponds to the amino acid sequence <SEQ ID 488; ORF58>:

```

1 .. LRETYAVLDS FDRYFVVALA GIFFFVRAQSE REWMRREVSAW QEKKGKEQKAE
51 LEPIKDGMFD FPELALMLFH AVKTAIVWLF VGUVVRFCRNY LAHESEPDRP
101 VPP..

```

Further work revealed the complete nucleotide sequence <SEQ ID 489>:

```

20 1 ATGTTTTGGA TAGTTTTGAT CGTTATTTTG TTGCTTGCCTT TTGCGGGCTT
51 51 GTTTTTGTC CGCGCACCAT CGGAACCGCA GTGATCGGC GAGGTTCTG
101 101 CGTGGCAGGG AAAGAAAGGG GAAACACGG CGGAGCTGCC TGAARTCAAA
151 151 GACGCTATGCC CGGATTTCG CGAACCTGCC CTGATGCTTT TCATGCCGT
201 201 CAAACCGGG CGTGTGATGCC TTGTTGTCGG TGTGCTCGGT TTCTGGCGGA
251 251 ACTATCTGGC CGCACGATTC CGAACGATGCC GACCCGGACCA GCGCGTCTTC
301 301 GCAACGGCTG CGGATGCTGC GACCGCATCC GACCGATATTAG CAGACAGTGG
351 351 AAACGGGAGC GAAAGAGCGG AAACCGGAGA AGCAGAAGCT CGGGAGGAAG
401 401 AGGCTGGCGA TACGGAGACG ATTGCAACTG CGCTTAATCGA CAACCGGCC
451 451 ATCCCAATTC ACCGGGATAT TTGCTGAGGG TTGATGCCGT CTGAAAGCGA
501 501 AATTGGCGCC GTCCGCGCGG TTGTTAAAGA AATCACTTG GAAGAAGCGAA
551 551 CGCGCTGTTT AAACCGCGCG GCTTAAAGGG AAACGAANAA ACGCTATATC
601 601 GATCTGATTG AGGAAAGACG AACACGGCGTC CCCAAAGTCG CGGTGTCGCGA
651 651 TACCGCGATG GAGGGCGCTGC AGATTTACGG TTGGAGACG CCTGTCGTTG
701 701 AACGCACGTA TTCCCATATG TTGATGCCGG ACACAAAGAC GTTTCCGAG
751 751 TCTGGGATTG ACGGTATTG GCGCTTATTT GAGAACGCGC ATCCGCTGTC
801 801 CTTTCTGCA GTCAAAAGCG AAAATCGACG GAAATCGCGC TTCCACCGTC
851 851 ATGAGGGGG GGGGAAAGGG CAGCGCGGGG CAAAAATCCCC GGATGTTTCC
901 901 CAAGGGCAGT CCTTTTACGAG CGCGACGCCG GTCCGCCGATC CCOCGCCCGG
951 951 CGTTCGCGC AATTGGAAAG AACCGACGACA GGCAACCGGGT TTCTCGGGAGG
1001 1001 CGCGAATTTC TCCTGCTGATT CGGGAAGTCG AGACGGTTGT CGGGAAACCG
1051 1051 GATCTGCGAA TGCGCTGCTG AGGAAAGAAAT GTTTTCACGG AAACCGTTTC
1101 1101 GTCTGGGGA CGCGCGCTGC CGGTTTATGA TGAAGACTGCC GATATCCATA
1151 1151 TTGAAAGAACCG TGCGCGCGGC GATCTCTGGG TTGTCGAACCC ACCCGGAATG
1201 1201 CGGAAAGTTG CCTATGCCGC AATCGATATT CGCCGCCGC CTCCCGTATC
1251 1251 GGAATATTCG AACCCTGACTC ATGACCGGCCG GTCAAGGATTC GAGCAGGTG
1301 1301 AACCGACGCCG CATTGCCGAC ACCGACCAATC TTGCCGATGA TTGTTTGAAAT
1351 1351 GGAGGTGGC AGGAGGAACAG CGCCGCTATAT CGCGATGACG GCAGTGAAGG
1401 1401 TCGCGGAGAGC CGGTCAAGCG GCGCAATATCT GTCGGAAACCC GAAGCGTTCG
1451 1451 CGGATGACAG TGACGGCTTG TTGCGGTTT AAAATGTCG GCTCTGAAACG
1501 1501 CGCTCTGCC GGGTATCGGA TACGGAGGGG ATGAGGGGG CGTCCCCATC
1551 1551 TGAAAGAACCG GTGGCGGTAT CGGAAACCTG CGCGACACCC GACCTGCTC
1601 1601 TGCTCTGCCG GTTCATCGAT CGGGCGACCC AAACCGAAAG AGAACCTGTTG
1651 1651 GAAAACAGCGA TCACCATCGA AGGAAATTAG CGCGAGTICA AGTCAAGGT
1701 1701 CAAGGGTTGGC GATCTTATT CGGGCGCCCGT AATTACGGT TATGAATCTG
1751 1751 AACCCGATGT CGCGCGGGC CGCAATTCG TTCTGAATCTG GAAAAAAAGAT
1801 1801 TTGGCGGGTT CGCTCGGGT GCCTTCCCATC CGGGTTGTG AAACCCATCCC
1851 1851 CGGGAAACCC TGCATGGTT TGGACATTCG GAACCGAAA CGCCAAATGTA

```

5	1901	TAGCGCTGAG CGAACATTCG AATTGGCCGC AGTTTGCGCA CTAACAAATC
	1951	AAGCTGCGC TCGGCTTCGG TCAGGACCG AGGGACGACG CGCTGTGAAAC
	2001	CGACTTGGGA AACGACCCGG ATTGTGTTGGF TTGGCGGACG AGGGCTTCGG
	2051	GCAAACTCGG GGGTGTCAA GCGATGATTC TGCTCATCTG TTCAAAAGCC
	2101	GCGCCGGAAG ACGTGCGTAT GATTTATGATC ATGCGGAAAAGA TTGCTGGATT
	2151	GAGCATTTACG GGAGAGGCG CCACACTGTG CCGCCCTGTG GTTACCGATA
	2201	TTGAGCTTGC CGAACAGCGG CTGAGGTTGG GTTAAAGCA GATGGAAAATC
	2251	CGCTACCGCC TGTAGATGGT TATGGGGGTG CGTAATCTTG CGGCCCTTCA
	2301	TCAAAAAAATC CGCCGAGCGG TGGAGAGGGG AAAAAGAAATC GCGAACATCG
	2351	TCAGCGCTCAC GCGCCGACGT CCCGAAACTT TGGAAAACAATC CGGCCCTTAT
10	2401	GGTGTGCTGG TCGTAGATGGT TGGCGGCTTC ATGATGAGCG CAGCGCAAGA
	2451	ATTAGCGAAGA CTGATGGTGG GCGCTGGCCA AAAGCCCGG CGGCCAGCG
	2501	TCCATTGATG TCTTGGCACA CAACCGCCCA GCGCTCGATG CATACCGGT
	2551	CIGATTAAGG GAAACATCC GACGCGTATC GGTTTCRAAG TGGTCAGCA
	2601	ATTACGACAGC CGCACCATC TGACCAAA GGGCGGGGAA AACCTCTCG
	2651	GTCAAGGGCA TAGTCTGGTG CTGGCGGCCG TAATGCTCTA CGGCCAGCG
	2701	TTTACGGCGC GTTGTGCGCTT GGATGAGAGG TGGTCAAAATA
	2751	TTTGAACAGC TTGGCGGAA CGCGATATGT TGCGATATTG TTGGTCGGG
15	2801	GGCGCGACGA AGACGTGCGG GCGCATCGGC CGAGCGGGA CGACGAAACC
	2851	GATCGCGATG AACGACGAGC CGTATCGCTT GTCGTTGAAAGA TTGCTGGGG
	2901	CAGCATTTGG CGCGTACAGC GGCGCTTGG CTATCGCTAC AACCGGGCGG
	2951	CGCGCTGTGAG TGGACGAGAT GAGGGCGGAAAG CGATTTGTCG CGACCGGGAA
	3001	CACAAAGCGA ACCTGGATAG CTGCTGCCCC TGTCGAACTG CTGCA

This corresponds to the amino acid sequence <SEQ ID 490: ORF58-1>:

1 MFWIVLVLIVL LLALAGLFFFV RAQSEREWMR EWSAQEEKK EKQAEELFEIK
 51 DGNPFDPELA MLLPHFKVTA VWTWFGVVR CRFNYLHAEZ EPDRVPVAS
 101 ANRADPTVATG DGSYDSEGGNT ESEAEEAEEA ESEAEEADTD IATAVIDNR
 151 IPFRSIAEG LMPSESEISV VRPVFKITL SEATRSLNS ALERTKKRY
 201 DAFEKNETCAT PVKRVSDUTM EGLQIQLGDD FVLRQTSYH FDADKEAFSE
 251 SADYGFEPYF EKHOHPFSVA KVAENARNAF THRHAGQSKQ AKASQFSDVS
 301 QGGSVSDCTA VDRARRRSVL NLKEPNKTT SAZARISRLS PESCVVGRK
 351 DVMPMSSETT VFTETTVSVS YGGVFYDETA DIHIEPAAP DAWVEPEVP
 401 PKVFMFTAID PPKVPPSVEIY NRTYPEPSSGF EOVRSRIAP TDHLADVLIN
 451 GGWOEETAAI DADGSEGAQ RSGSYGLSET EAFGHDOSAV CFVNPFSEV
 501 PCSRUDSTDEA DEAGFPSEET GAVSHLBPTT DDLPLPFLNEV EATOTEEELL
 551 ENSTIIEKLU AEFVKVFKWV DSYSGVPITR YEIEPDVGVU GNSVNLAKED
 601 LARSIGVASI RVETETV:PGK CMGLLEPNPK RQMRILSEIF NSPEFAEKS
 651 KTLZLAGOD: TQGPVFTTDLA KAPHILVAGT TSGGSKVGVN AMILSMLKA
 701 APEDVRMVIMI DPKMLHELSIY EGIPHILVAP VTMDKLAAN LNWCWNEMEK
 751 RYRIMSPGMV RLNAGFNGKQ AAAEARGEKI GNFSPLTPD PEPLXPLKFI
 801 VVVVDFEADL MMTAGKXKIE LIARLAQKQ AGHLILHAT QRPSWDVING
 851 LIKANIPTRE AFQVSSKDS RTILDQMGAE NLLGQGDMLF ILPCTAYOR
 901 VHGFASDEEV VRVHRVEYLIQK FGEPEVDYI LSSGGSEELP GIGRSQGDET
 951 DPMYDEAVS VKLTRKASIS GVQRALRIGY NRAARLIDQW EAECIVSAE
 1001 HNGNBTJLVP LDNA*

Computer analysis of this amino acid sequence predicts the indicated transmembrane region, and also gave the following results:

Homology with a predicted ORF from *N. meningitidis* (strain A)

QRE58 shows 96.6% identity over a 89aa overlap with an QRE (QRE58a) from strain A of *N.*

50 *meningitidis*:

		10	20	30	40	50	60
	orf58.pep	LRETAYVLDSFDRYFVVALAGLFFVRAQSEREWRMREVSQAWQEKGKEQKQAELEPIKDGMFD					
	orf58a	<u>MFWIVLIVLIVLALAGLFFVRAQSEREWRMREVSQAWQEKGKEQKQAELEPIKDGMFD</u>					
55		10	20	30	40	50	
	orf58.pep	FPELAALMFHAVKTTAVYWLFGVGVRFCRNYLIAHESEPDRPVPP					
	orf58a	<u>FPELAALMFHAVKTTAVYWLFGVGVRFCRNYLIAHESEPDRPVPPASANRADVTASDGYSD</u>					
60		60	70	80	90	100	110

The complete length ORF58a nucleotide sequence <SEQ ID 491> is:

```

      1 ATGTTTTGGA TAGTTTGTAT CGTATTTCG TTGCTTGGCGC TTGCCGGCTT
      5 GTTTTTGTC CGGCCACCAT CGAACCGGCR GTGGATGGCC GAGGTTCCTG
     10 CCGGGCAGG AAGAARAGGG GAAACACGG CGGAGCTCGG TGAARTCAAA
     15 151 GAGCTATGC CGGATTTCGCG CGAACATTCG CTGATGCTT TCCATGCCGT
     20 201 CAAAACCGCA GTGATTGGC TTGTTGTCGG TTGCTGCCGT TTCTGCCGAA
     25 251 ACTATCTGGC SCACGATCC GAACCGGACA GCCCCGCTTC GCTGCTTCT
     30 301 CGAAATCGTG CGGATGTTCC GACCCGATCC GACGGATATT CAGACAGTGG
     35 351 AAAACGGGACG GAAGAAGCGG AAACGGAAGA AGCAGAGCT GCGGAGGAAG
     40 401 AGGCTGCGCA TACGGAGAG ATGCAACTG CGCTAATCGA CAACCGCCGC
     45 451 ATTCACATTCC ACCGGATATG TGCTGAAGGG TTGATGCCGT CTGAAAGCGA
     50 501 AATTTCGCCG CTCGGCTCG TTGTTAAGGA ATTCATTTTG GAAAGAGCAA
     55 551 CGCGTGCCTT AAACAGCGGC GCTTTAAAGGG AAACGAAAAA ACCTATATC
     60 601 GATGATTTG AGAAAAGACG AAACAGCGGT CCCAAAGCTG GCGTGTCCGA
     65 651 TACCCCGATG GAGGGCTGC AGATATTCGG TTGGAACGAG CCTGTCCTC
     70 701 AACGACGCTA TTTCCCATAG TTGATGTCGG ACAAAGAACG TTHTTCCGAG
     75 751 TCTGGGATTG CGGATTTCGA CGCTTATTTT GAGAACGAGC ATCCGCTTGC
     80 801 CTTTTCCTGC CGTAAACGGC AAATAGCAGG GAAATGGCCG TTCCGCCGTC
     85 851 ATGAGGCGGC GGGNAAGGG CAGGGAGGGG CNAATCCCC GGATTTTCCC
     90 901 CAAAGGCGATG CGGTTCAGA CGGCACAGCG TTCCGGGATC CGNGCCGCCG
     95 951 CGTTTCCGTC AATTTGAAGA AACCAGAACN GCAACCGGT TTCTGGGAGG
    100 1001 CGCGGATTTC CGGCCTGATT CGGGAAGCTG GGACGGTTGT CGGGAACCG
    105 1051 GATGTCGAAAG TGGCTCTGA AACCAGAAA GTTTTCAGG ARANTGTTTC
    110 1101 GTCTGTGGGA TACGGCGNTC CGGTITATGA TGAAACTGCG GATATCCATA
    115 1151 TTGAAGAACCC TTGGCGGCCG GTGCTTGGG TTGTCGAACCC ACCGGAAGTG
    120 1201 CGGAGATGTC CATGGCGCC ATTGNATATT CGCGCGCCCG CTCCGATATC
    125 1251 GGAATCTAC AACCCTACCT ATGAAACGGC GCGAGGATTC GACAGGATTC
    130 1301 AACGACAGCG CATTGGCCA ACCGATCATO TTGCGATGA TTGTTTGAAT
    135 1351 GGAGGTTGGC AGGAGGAAGA CGCCGCTTAT CGGAATGACG GCAATGAGG
    140 1401 TGTCGAGAG CGTCGAGGG GCGAATTTTG TTGCGAACG GAAAGCTTCG
    145 1451 GGCATGACAG TCAGGGGGT GTGCGTTTG AAAATGTGCG GTCTGAACGC
    150 1501 CGCTCCCGCC GGGCATNGGA TACGGAAAGG GATGAAGGGG CTTCACATC
    155 1551 TGAAGAACCC GTGCTGGAT CGGACACCTT CGCGACACCC GACCGCTTTC
    160 1601 TCGCCGGCGT GTTCAATCC CGGGAGGAGC AAACCGAAGA AGANCTGTTG
    165 1651 GANACACAGA TCACCATCGA AAAAATGAT CGGGAGTTCA AAAGTCAAGGT
    170 1701 CAAGGTGGC GATTTCTTAT CGGGCCCGT GATTAGCGG TATGAAATCG
    175 1751 AACCCGATG EGGCGTGGG CGCAATTCGG TTCTAAATCT GTAAAAGAGAN
    180 1801 TTGGCGGTT CGCTCGGGG GCGCTTCATC CGCTTGTGC AAACCATCT
    185 1851 CGGCNAAAACCG TTGATGGTGT TTGAAACTCTT GAAACCGGAA CGCCAAATG
    190 1901 TACGCTCTGAG CGAACATTCG AATTGCGCCG AGTTGCGGA ATCCAAATCC
    195 1951 AAAGCTGCCG TCGCCGTCG ICAGGACATC CGCGACACG CCCTCTAAC
    200 2001 CGACTTGGCC AAACGCCCG ATTGTTGCGT TGCCGCCAGC ACCTGTTGG
    205 2051 GCAATTGGT EGGTGTCAAC CGGATGATTC TGCTATGCT TTTCAAAGCC
    210 2101 GGGCGGAAGG ACCTGTTGAT GATTATGATC GATCGGAAAAT TCTGGAATT
    215 2151 GAGCATTTAC GGACGATCC CGCACCTGCT CGCCCTGTG GTTACCGATA
    220 2201 TGAAGCTGGC CGCAACACGG CTGAACTGGT GTGTTAACCGG AATGGAAAAAA
    225 2251 CGCTACCGGG TGTAGGACCTG TATGGCGCG CGCAATCTT CCGGNTCAA
    230 2301 TCAAAAATAC CGCGAACGCC CGAACAGGGG CGAGAAAATG GCAACCCGT
    235 2351 TACGGCTTCAC CGGCCGACAT CCCGACATT TGGANAATT GCGGTATTAC
    240 2401 GTGCTGGTGG TTGATGAGTC TCCGGACCTT ATGATGACCG CAGCGAAGAA
    245 2451 ATTCAGAAGA CTGATGGCCG GCCTGGCCCA AAAACCGCCG CGCGCAGGCC
    250 2501 TCCCATTTAT CGGACGACCA CGAACGCCCA GTGTCGATGT CATCACCGGT
    255 2551 CTGATTAAGG CGAACATCCG GACGGGTATC CGCTTCACAG TGTCAGCGA
    260 2601 ATTCAGAACGC CGAACGGCTC TTGACCAAAT GGGTGGGAA AACCTGCTCG
    265 2651 CGGAGGGCGA TATGCTGTC CGCGGGCGG GTACGGCTTA TCCGAGCGC
    270 2701 GTTCACCGGG CGTTTGCCTC GGATGAGAG CGGACACCGG TTGTCGRATA
    275 2751 TCTGAAACAGC TTGCGCGAC CGGACTATGT TGCGATATN TTGAGCGGGC
    280 2801 GTATGTCGGA CGATTGGTGG GGAATCACGG CGACGGCGCA CGCGGAACCC
    285 2851 GATCCGATG ACAGCAGGGC CGTGTGCGTT TTGTTGAAAA CGCGCAAGAC
    290 2901 CAGCATTTCT CGGCTGCGAG CGCCATIGCG TATCGCTAT AATCGCCGCG
    295 2951 CGCGCTCTGAT TGACGATGAG GAGCGGAAGG GCAATTGTTG CGACCGGAA
    300 3001 CACACCGGCA ACCGATGACT TCTGTCCCCC TTGACAAATG TTGTA

```

This encodes a protein having amino acid sequence <SEQ ID 492>:

```

      1 MFWIVLILIV LLALAGLFFV RAQSEREWRM EVSAWQEKKK EKQAEELPEIK
      5 DGMGPDPFLA LMLHEVKITA VYWLVEFGVR FCRNLYLAHES EPDRPVPPAS
     10 101 ANRADVPTAS DGYSDSGNGT EEAETEEAEE AEEAAADTED IATAVIDNRR
     15 151 IPEFDRSIAEG LMPSESEISI P VRVEKEITL EEAATRALNSA ALRETKKRYI
     20 201 DAFEKNETAV PKVVRVSDTPM EGLQIIGLDD FVLRQTSRM FDADKEAFSE

```

251 SADYGFEPYF EKQHPSAFSA VKAENARNAP FRRHAGCQKG QAEAKSPDVS
 301 QGQSVDGTA VRDAXRVRSV NLKEPNKATV SAEARISRLI PESRTVVGKR
 351 DVEMPSETEN VFTEXXSSVG YKXPVYDETA DIHIEEPAPW DWAMWVPEPEV
 401 PKVMPMAKDI PPPPPVSEIY NRTYEPPEAGF EQVQRSLIAE TDHLADDVLN
 451 GGWEETAAI ANDCSEGVAR RSSCQYLSST EAEGHDOSAV CFFENVPSE
 501 PSRAXDTEA DECAFQSEET GAVSEHLPPT DLLLPFLFNP GATQTEEXLL
 551 XNSITIEEKX AERPKVKVVA DSYSGPVIR YEIEPDVGVE GNSVINLEKX
 601 LARSLGVASI RVVETLICKY CMGELPNPK RQNIRLSEIF NSPEFAEKS
 651 KLTIALCQDI TGCPVVTDLG KAPHLLVACT TGCKSGVGN ANILSMLFLKA
 701 APEDVRMIMI DPXMLELIV EGPILLALV VTDMKLAANNA LNWCVNEMEK
 751 YRILMSFMGV RNLAGXNKX AEAAAERKEI CNFSLTBDN PEPLXKLFFI
 801 VVVDEEADL MMTAGKKIEE LIARLQQKAR AACIHLILAT QRPSPVDVITG
 851 LIKANIPTRI AFUVSSKIDS RTILDGMAGS NLGGCGDMLF LPEGTAYPOR
 901 VHGFASDDEE VHRVVEYLQK FGEPDVDDX LSGGMSDDLL GISRSGDGET
 951 DPMYDRAEVSV VLTKRKASIS GVQRALRICY NRAARLIDQM EAEGIVSAPE
 1001 HNGNRTILVP XDNA*

ORF58a and ORF58-1 show 96.6% identity in 1014 aa overlap:

	10	20	30	40	50	60
20 orf58a.pep	MFWIVLIVLILLALAGL	FVRAQSEREWMRREVSAWQ	SEKKGEKQAEELPEIKDGM	PDPFELA		
orf58-1	MFWIVLIVLILLALAGL	FVRAQSEREWMRREVSAWQ	SEKKGEKQAEELPEIKDGM	PDPFELA		
	10	20	30	40	50	60
25 orf58a.pep	70	80	90	100	110	120
orf58-1	LMLFHAVKTAVYWLFVG	VVRFCRNLYLAHESEPD	RPVPPASANRADVPTASD	GYSDSGN	GT	
	70	80	90	100	110	120
30 orf58a.pep	130	140	150	160	170	180
orf58-1	EAEATZEEAEEAEEA	ADDETDIATAVIDNR	IIFPDRSIAEC	CLMPSSSEIISP	VPVRVPKEITL	
	130	140	150	160	170	180
35 orf58a.pep	190	200	210	220	230	240
orf58-1	EEATRALNSAALRET	KKRYIDAFEKN	ETAVPKVVRVS	DTPMEGLQIIGL	DDPVVLQRTYS	GT
	190	200	210	220	230	240
40 orf58a.pep	250	260	270	280	290	300
orf58-1	FDADKEAFSESADYG	FEFYFEKQHPSAFSA	VAKENARNAP	FRRHAGCQKQ	QAEAKSPDVS	
	250	260	270	280	290	300
45 orf58a.pep	310	320	330	340	350	360
orf58-1	QQGSVSDGTAVR	DARRSVNLKEPNKATV	SABARI	SRLIPESRIVV	GKRDVEM	PSSETEN
	310	320	330	340	350	360
50 orf58a.pep	370	380	390	400	410	420
orf58-1	VFTETVSSVG	YGGCPVYDETADIHIE	EPAPADAWV	VEPPEVKV	PMAPDIPPPP	PVSEIY
	370	380	390	400	410	420
55 orf58a.pep	430	440	450	460	470	480
orf58-1	NRTYEPPEAGF	EQVQRSLIAETD	HLDADDVLNGW	QEETAAI	ANDCSEGVAR	SSGQYI
	430	440	450	460	470	480
60 orf58a.pep	490	500	510	520	530	540
orf58-1	EAEGHDOSAQ	VCPFENVP	SERPSRAXDTE	TADEGAFQ	SEETGAVSEH	LPTTDLLLP
	490	500	510	520	530	540
65 orf58a.pep	590	600	610	620	630	640
orf58-1	EAEGHDOSAQ	VCPFENVP	SERPSRCSRV	DTADEGAF	QSEETGAVSEH	LPTTDLLLP
	590	600	610	620	630	640
70 orf58a.pep	690	700	710	720	730	740
orf58-1	EAEGHDOSAQ	VCPFENVP	SERPSRCSRV	DTADEGAF	QSEETGAVSEH	LPTTDLLLP
	690	700	710	720	730	740

-290-

		550	560	570	580	590	600
	orf58a.pep	GATQTEEXLLNSITIESKKAEFPKVVKVVDSYSGPVITRYIEPDVGVRGNCSVNLKEK					
5	orf58-1	ATQTEELLENESITIESKKLAEFPKVVKVVDSYSGPVITRYIEPDVGVRGNCSVNLKEKD					
		550	560	570	580	590	600
	orf58a.pep	610	620	630	640	650	660
10	orf58-1	LARSLGVASIRVVEITILGKTCMGLELPNPKRQMIRLSEIIFNSPEFAESKSKLTLALGODI					
	orf58a.pep	LARSLGVASIRVVEITILGKTCMGLELPNPKRQMIRLSEIIFNSPEFAESKSKLTLALGODI					
	orf58-1	610	620	630	640	650	660
		670	680	690	700	710	720
15	orf58a.pep	TGQPVVTDLGKAPHILLVAGTTGGSKSGVNVAMILSMSLEKKAPEPDVRIMIDPKMLELSIY					
	orf58-1	TGQPVVTDLGKAPHILLVAGTTGGSKSGVNVAMILSMSLEKKAPEPDVRIMIDPKMLELSIY					
	orf58-1	670	680	690	700	710	720
		730	740	750	760	770	780
20	orf58a.pep	EGIPHLAPAVTDMKLLAANALNCWVNEMEKRYLRLMSFMGVRNLAGXNQKIAAAARGEKI					
	orf58-1	EGIPHLAPAVTDMKLLAANALNCWVNEMEKRYLRLMSFMGVRNLAGXNQKIAAAARGEKI					
	orf58-1	730	740	750	760	770	780
		790	800	810	820	830	840
25	orf58a.pep	GNPFSLTDPNPEPLXKLFPIVVVVFDEFADLMMTAGKKIEELIARLAQKARAAGIHLILAT					
	orf58-1	GNPFSLTDPNPEPLXKLFPIVVVVFDEFADLMMTAGKKIEELIARLAQKARAAGIHLILAT					
	orf58-1	790	800	810	820	830	840
		850	860	870	880	890	900
30	orf58a.pep	QRPSVDVITGLIKANIPTRIAFOVSSKIDSRSTILDQMGAEPLLGGDMFLPPGTAYQR					
	orf58-1	QRPSVDVITGLIKANIPTRIAFOVSSKIDSRSTILDQMGAEPLLGGDMFLPPGTAYQR					
	orf58-1	850	860	870	880	890	900
		910	920	930	940	950	960
40	orf58a.pep	VHGFASADEEVHRVVEYLKQFGEPDVDDXLSSGMSSDDELLGISRSRGDDETDPMYDEAVSV					
	orf58-1	VHGFASADEEVHRVVEYLKQFGEPDVDDXLSSGMSSDDELLGISRSRGDDETDPMYDEAVSV					
	orf58-1	910	920	930	940	950	960
		970	980	990	1000	1010	
45	orf58a.pep	VLKTRKASIISGVQLARIIGYNRAARLIDQMEAEGIVSAPEHNGNRTILVPLDNAX					
	orf58-1	VLKTRKASIISGVQLARIIGYNRAARLIDQMEAEGIVSAPEHNGNRTILVPLDNAX					
	orf58-1	970	980	990	1000	1010	
		1000	1010	1020	1030	1040	
50	Homology with a predicted ORF from <i>N.gonorrhoeae</i>						
	ORF58 shows complete identity over a 9aa overlap with a predicted ORF (ORF58ng) from <i>N.gonorrhoeae</i> :						
	orf58a.pep	ALMLPHAVKTAVYWLFGVGVRCNRNLAHESEPDORPVPP					103
55	orf58ng	SEPDRPVPPFASANRADVPTASDGYSDSGNG					30

The ORF58ng nucleotide sequence <SEQ ID 493> is predicted to encode a protein having partial amino acid sequence <SEQ ID 494>:

1	..SEPDPRUPPA SANRADWPTA SDGYSDSGNG TEEAEETEAE AREEEARDTE					
51	51 DATAVIDNR RIFTDSIAE GLMQSESKTS PVRPVKEKIT LEEATRALSS					
101	101 AALRETDKRYY 1DAFKEKNTGA VPKVVRSDTP MEGLQIIGLD DPVLORTYSR					
151	151 MFDADKECS ESADYGPEPY FEKOHPSPSF AVKAENARNA PFRRHQAGQEK					
201	201 GQAEEKSPDV SQGQSVDGT AVRDAARRVS VNLEKPNKAT VSAEARISR					
251	251 IPESRTVVGG RDVEMPSETE NVFTETVSSV GYGGFPVYDEA ADIHIEEPAA					
301	301 PDAWWVEPPE VPEVAVPEID ILPPPVPSEI YNRTYEPPAG FEQAQRSRIA					

351 ETDHLLAADV L NGGWQEQETAA I ADDGSEGAA ERSSGQYLS TEAFGHDSQA
 401 VCPFDVPSFE RPSCRVSDTE ADEGAFQSEE TGAVSEHLEPT TDLLLPFLFN
 451 PEATQTTEEL LENSITIEEK LAEFKVKRVK VDSSYSGFVIT RYEIEPDVG
 501 RGNSVLNLKEK DLARSIGVAS IRVETTIPKG TCMGLELFPN KRMIRLSEL
 551 FNSPEFAESK SKLTALGGD ITGQPVFTDL GKAPHILLVAG TTGSGKSVGV
 601 NAMILSMLFK AAPEFDVRMIM IDPKMLELSI YEGITHLLAP VVTDMKLLAN
 651 ALNWCVNEME KRYKRLMSFMA VRNLAGFNQK TAEAAARGEK IGNPFSLTP
 701 DEEPEPLEKLPT^{IVVVVDEFAD} LMWTAGKKIE ELIARLAQKA RAAGIHLLA
 751 TCRPSWDVIT GLIKANIPTR IAFQVSSKID SRTILDQCMGA ENLLQGQDML
 801 FLPGPTAYPQ RVHGAFASDE EVHRVVEYKL QRGEPPDYVDD ILGGGGSEEL
 851 PGIGRSGDGE TDPMYDEAVS VULVTRKASI SGVQRALRIG YNRAARLIDQ
 901 MEAEIGVSAF EHNGNRTILV PLDNA*

This partial gonococcal sequence contains a predicted transmembrane region and a predicted ATP/GTP-binding site motif A (P-loop; double underlined). Furthermore, it has a domain homologous to the FTSK cell division protein of *E. coli*. Alignment of ORF58ng and FtsK (accession number p46889) show a 65 % amino acid identity in 459 overlap:

ORF58ng:	467	IIEKLAEEFKVKVKVVDYSYSGPVITRYIEIPDVGVGRGSNVINLEKDLARSLIGVASIRVVT	526	
	+E +LA+F+K	VV+ GIVTR-E+ G+ NL +DLARSL	++RVVE	
FtsK:	868	VEARLADFKIADKVNNYSFGFVITREFLNLPVGKAANISLNSRDLARSLSTVAVRVRVEV	927	
20	ORF58ng:	527	IIGKTCMGLELENPKRQMTRLSEPEAESKSKLTLALGGDITGQPVFTDLGKPHL	586
	IGK +GLEEN	KRC + L E+ ++ F+ S LT+ LG+D G+PVV DL K PHL		
	FtsK:	928	IPGKPYVGLLELPNKKRQTVYLREVLDNAKFRDRNNSPLTVLIGKDIAGEPVVADLAKMPHL	987
25	ORF58ng:	587	LVAGTTGSGKSVGVNAMILSMFKAPADEPVNMIMIDPKMELSLIYEGITHLLAPVVTDMK	646
	LVAGTTGSGKSVGVNAMILSM+KA	PEDVA IMIDPKMELSL+YEGI HLL VVTDMK		
	FtsK:	988	LVAGTTGSGKSVGVNAMILSMLYKQAPEDVRFIMIDPKMELSLSYEGIPHLLTEVVTDMK	1047
30	ORF58ng:	647	LAANALWCNCNEEMKRYRLMSFMSVNVNLAQGPNKQKIAAARARGEKIGNPFSLTPDPER--	704
	ANAL WCNCNE+RV+LMS	+GVNRNLAG+N+KIAEA		
	FtsK:	1048	DAANALWCNCNEMERRYKUMLSALGVNRNLAG+N+KIAEADRMRRPIPDPPWKGPMDSMDAQH	1107
35	ORF58ng:	705	--LEKLPFIVVVVVFEDALIMMTAGKIEELIARLAQKARAAGIHLILATQRSPVDVITGL	762
	L+P +IV+V-VDEFADLMMT	GKK+EEELIARLAQKARAAGIHL+LATQRSPVDVITGL		
	FtsK:	1108	PVLKKEPIVVLVDEPFDALIMMTVGKVKVEELIARLAQKARAAGIHLVLATQRSPVDVITGL	1167
40	ORF58ng:	763	IKANIPTRIAFVSSKIDSRTILDQMGAEANLLGGGMLFLPPTAYPQRVHGAFASDEEV	822
	IKANIPTRIAF VSSKIDSRTILDQ GAE+LLG GDML+	P + P RVHGAF D+EV		
	FtsK:	1168	IKANIPTRIAFTVVLSKIDSRTILDQAGAATLQGMSLGMGMLYSGPNSTLPVPRVAFVARDQEY	1227
45	ORF58ng:	823	HRVVEYLKQFGEPDYVDDILSGGGSSEELPGIGRSGDGETDPMYEAVSVVLKTRKASISG	882
	H VV+ K G P YVD I S SE G G E D+E+D+AV V + RKASISG			
	FtsK:	1228	HAVVQDWKARGRPQYVGDITDSSEGGAG-GFDGAEEELDPLFDQAVQFVTEKRKASISG	1286
50	ORF58ng:	883	VQRALRIGYNNRARRLQMEAEIGVUSPHEHNNTILV	921
	VQR RIGYNNRARR+I+QMEA+GIVS	HNGR +L F		
	FtsK:	1287	VQRQFRIGYNNRARRIIEQMEAQGIVSEQHGNHNRREVLF	1325

Further work on ORF58ng revealed the complete gonococcal DNA sequence to be <SEQ ID 495>:

50	1	ATGTTTTGGA TAGTTTTGTG CGTTAT:ggt	TTGCTTGCGC TTGCGGGCCT
	51	GTTTTTGTG CCGGCACCAT CGGAACCGGA GTGGATGGC GAGGTTTCG	
	101	CGTGGCAGGA AAAGAACGGG AAAAACACGG CGGAGCTGCC	TGAATCAAA
	151	GACGTTATGC CCGATTTCG CGAGTTTCC CTGATGCTTT	TCCATCGGT
	201	CAAAACCGGA GTGATATTGG TGTGTTGCTGG TGTGCTGCCG	TTCTGCGGAA
	251	ACTATCTGGC GCACGATCTC GRACCCGACCA GGCCCCCTTC	GCGCTCTCT
	301	GCAAAACCGT GGATGATTCG GACCGCATTC GACCGGGTTAT	CAGACAGTGG
	351	AAACGGGAGC GAAGAACGGG AAACGGGAGC AGCACAGCT	GGCGGAGGAG
	401	ACGCTGCCA TACgGAAGAC ATTGCGAACG CGCTTAATGGA	CAACCGCCGC
	451	ATCCGATCTG ACGGCGATAT TGCTGAAGG TTGATGCACT	CTGAAAGCAA
	501	AACCTGGCCC GTCCGTCGGG TTTTTAAAGGA ATCACTTTG	GAAGAAGCAA
	551	CGCGTGTCTT AAGCGACGGG GCTTTAAAGGG AAACGAAAAAA	ACGCTATATC
	601	GATGCGATTG AAAAACACGG AACACCGCGT CCCAAAGTAC	GGGTGTCGGA
	651	TACCCCGATG GAAGGCGCTGC AGATTTATCGG TTGTCAGCAG	CTCTGCTTC
	701	AAACGCACTGA TTCCCGATAG TTGATGCGG ACAAGAACGC	GTTCGCGAG
	751	TCTGGGATT ACGGATTTGA GCGCTATTG GAGAACGAGC	ATCCGCTGTC

	801	CTTTTCTGC	GTCAAAGCCG	AAAATGCCAG	GAATGCGCCG	TTCGGCCGTC
	851	ATGAGGGCA	GGAGAAAGGG	CAGCGGGAGG	CAAAATCCC	GGATGTTTCC
	901	CAAGGCGCATG	CGGTTCAGA	CGGCACAGCG	GTCGGCGATG	CGCGCGCCG
5	951	CGTTTCGGTC	AATTGGAAG	AACCGAACAA	GCCAACCGTT	TCTGCGGAGG
	1001	CGCGGATTC	GGCGCTGATT	CGGGAAAGTC	GGACCGTTGT	CGGGAAACCG
	1051	GATGTCGAA	TGGCGCTCTGA	AACCGAAANT	GTTTTCACGG	AAACCGTTTC
	1101	GTCCTGGGG	TACGCCGGTC	CGGTTTATGA	TGAAGCTGCC	GATATCCATA
	1151	TTGAAGAGCC	TGGCGCCGCC	GATGCTTTGG	TGGTGAACCC	ACCCGAAGTG
10	1201	CGGGAGGTAC	CGTAGCCCGA	AATCGATATT	CTGGCGCCG	CTCCCGTATC
	1251	GGAAATCTAC	AACCTACACT	ATGAGGCCGC	GGCAAGGATTG	GAGCAGGCCG
	1301	AACCCAGCCG	CATTGCCGAA	ACCGCACCATC	TTGCCGCTGA	TGTTTTGAAT
	1351	GGAGGTGGTG	AGGAGGAAC	CGCCGGTATT	CGACATGACG	GCAGTGAAGGG
	1401	TSGCCGAGAG	CGGCTAACGG	GGCAATATTC	GTGAAAACC	GAAGCGTTCG
15	1451	GGCAGTACAC	TCAAGGGGTT	TGTCCTTTTG	AAAGATGTCG	GTCTGAAACG
	1501	CGCTCTGGC	GGGTATCGGA	TACCGAAGGG	GATGAAAGGG	CTTCCAAATC
	1551	GGGAAGGACG	GTTGCGGTAT	CGGAACACCT	GCGACAAAC	GACCTGCTTC
	1601	TGCTCCGGC	GTTCATCCC	GAGGCCAGCG	AAACCGAAGA	AAACACTGTTG
	1651	GAANACAGCA	TCAACATCA	AAAAAAATTC	GCGGAGTTCA	AAGTCAGGTT
20	1701	CAAGGTGTC	GATCTTATT	CGGGCCCCGT	GATTACCGGT	TATGAAATCG
	1751	AACCCGATGT	CGGCCGCGGC	GGCAATTCTCG	TTCGAATTTC	GGAAAAAACG
	1801	TTGGCGGTG	CGCTGGCGGT	GGCTTCCATC	GGGTTTGCG	AAACCATCTCC
	1851	CGGAAAACC	TCACTGGTT	TGGAACCTTC	GAACCCGAAA	CGCCAAATG
	1901	TACCGCTG	CGGAATTC	AATTGCCCCG	AGTTTGGCGA	ATCCAAATCC
25	1951	AAGCTGACG	TGGCGCTCGG	TCAAGCAGAT	ACGGGACAGC	CGCGCTAAC
	2001	CGACTTGGG	AAAGCACCGC	ATTGGCTGGT	TGCGCGCACG	ACCGGTTCCG
	2051	CGAAATCGGT	GGGTGTCAAC	CGGATGATTG	TGTCATGCT	TTCAAAAGCC
	2101	GGCGCGGAAG	ACGTCGCTAT	GATTATGATC	GATCCGAAAAA	TGCTGGAAATT
	2151	GAGCATTAC	GGAGCATCA	CGCACCTGTC	CGCCCGTGT	TTACCGATA
30	2201	TGAGCTGGT	GGCAAAACGGC	TGCAACTGGT	GTGTTAACGA	AATGGAAAAAA
	2251	CGCTACCGCC	TGATGAGCTT	TATGGCGGT	CGCAATCTTG	CGGCGCTTCAA
	2301	CCAAAAAATC	GGCCGAGGCC	CAGCAAGGGG	AGAAAAAAATC	GGCAATCTCGT
	2351	TGACCTCTAC	GGCCGAGCAT	CGCCGACCT	TGAAAAAAACT	CGCTTTTATC
	2401	GTGTCGTTG	TGATGAGTT	TGCGGTTTG	ATGATGACG	CAGGCAAGAA
35	2451	AATGAGGAA	CTGATGGCG	GGCTGCCCGA	AAAGGCCGC	GGCGCAGCGA
	2501	TCCACCTTAT	CTTGGCCACA	AAAGGCCCGA	GGCTGCGATG	CATCACCGGT
	2551	CTGATTAAGC	CGAACATC	GAGCGGTATC	GGCTTCCAAAG	TGTCAGCAA
	2601	AATGAGACAGC	CGCACGATTC	TGACCAAAAT	GGGGCGGAA	AACTCTGCG
	2651	GTGACGGCGA	TATGCTGTC	CTGCGCGCG	GTACTGCGTA	TCCCGACGCG
40	2701	GTTCACGGCG	CGTTGGCG	GGATGAGAG	GTGACCGCCG	TGTCGGAATA
	2751	TCTGAGAACG	TTTGGCGAGG	CGGACTATGT	TGACGATATT	TTAGGGGGGG
	2801	GGCGCAGCGG	AGACTTCCC	GGCATCGGG	GCAGGGCGA	CGCGGAAACC
	2851	GATCGATGAT	ACGAGGAGG	GTATTCGGT	GTCTGAAAAA	CGGCCAAAGC
	2901	CAGCATTTC	GGCGTACAGC	GGCGCTTGGG	CATCGCTAC	AAACCGGGCG
45	2951	CGCGCTGAT	TGACCAAATG	GAAGCGGAAG	GCATTTGTGTC	CGCACCGGAA
	3001	CACACCGCGA	ACCGTCAGAT	TCTCGTCCCC	TTGGACAAATC	CTTGA

This corresponds to the amino acid sequence <SEQ ID 496; ORF58ng-1>:

	1	MWIVLIVIV	LILAGLFLV	RQSEREWIR	EVSAMQEKKG	EKQAEPEI
	51	DGMDFPFES	IMLFHAYKTA	VVILFVGUVR	FCRNYLAHES	FPDRVPVPPAS
	101	ANRADPTAS	DGYSIDSGNGT	EEAEETTEAAA	EEAAEAUTED	IATAVIDNRR
50	151	IPFDRIABE	IMQSESKTSE	VRPVFKIEITL	EEATRALSSA	ALRETKRKYI
	201	DAFKNGTAT	EGSLSKTSV	EGLQIIGLID	PVLQRYSRMR	FDADKEAFSE
	251	SADYGFEPY	EKQHPSAFSA	VKAENARNAP	FRRHAGQEQKG	QAEAKSPDVS
	301	QGGSVSDGTA	VRDARRVSV	NLKEPNKATV	SAEARISRLI	PESTRTVVGKR
	351	DVEMPSETEN	VFTETVSSVG	YGGFVDEA	DIHIEEPAAF	DAWVEPEPEV
	401	PEVAVPEIDI	LFFFPSSEIY	NRTYEPPAGF	EQAQRSRIAE	TDHLAADVLN
	451	GGWQEETAAI	ADDGSEGAAE	RS5QYLVSET	EAFGHDQSACV	CPFEDVPSER
	501	PSCRVSDTEA	DEGAFQSEET	GAVSEHLLPTT	DLLLEPLFLNE	EATOTEEELL
	551	ENSTIEEKL	A5FKVKVVK	DSYSGFVITR	YEIEPDVGVR	GN5VNLNEKD
	601	LARSLGVASI	RVVTETLPKT	CMGLELPBNPK	RQMIRLSEIF	NSPEFAEKS
60	651	KLTIALQGD	TQGPVVDLG	KAPHLLVAGT	TGSGKSVGVN	AM1SLMFLKA
	701	APEDVRMIM	DKMLLESIY	EGITHLLALP	VTDMKLAANA	LNWCVNEMEK
	751	YRILMSFMGV	RNLAGNQNKI	EEAAAERGEKI	GNFSLTPDD	PEPLEKLPI
	801	VVVUDFEDAL	MMTAGKIKEF	LIRLAAQKAR	AAGIHLLAT	QRB5VDOIING
	851	LIXKANIPTRI	AQFVQSSKIDS	RTILDQMGE	NLLQQGDMLF	LFPCTAYPQR
	901	VHGAFASDEE	VHRVUEYLQ	FGEPDVDDI	LGGGSEELP	GIGRSGDGET
	951	DPMYDEAVPS	VLK7KASIS	GVQRALRIGY	NRAARLIDQM	EAEGIVSAPE
65	1001	HNGNRTILVP	LDNA*			

ORF58ng-1 and ORF58-1 show 97.2% identity in 1014 aa overlap:

		10	20	30	40	50	60
	orf58-1.pep	MFWIVLIVILLALAGLFFVRAQSEREMRREVSAWQEKKGKEKQAEELPEIKDGMPDFPEFLA					
5	orf58ng-1	: : : : : : : : : : : : : MFWIVLIVILLALAGLFFVRAQSEREMRREVSAWQEKKGKEKQAEELPEIKDGMPDFPEFLA	10	20	30	40	50
							60
		70	80	90	100	110	120
	orf58-1.pep	LMLFHAVKTAJVYWLFGVVRFCRNLYLAHESEPDRPVPPASANRADVPTASDGYSDSGNGT					
10	orf58ng-1	: : : : : : : : : : : : LMLFHAVKTAJVYWLFGVVRFCRNLYLAHESEPDRPVPPASANRADVPTASDGYSDSGNGT	70	80	90	100	110
							120
		130	140	150	160	170	180
15	orf58-1.pep	EEAETEEAAEAEAAAADTEDIATAVIDNRRIIPFDRSIAECLMPSSESEISFSPVRVPKEITL					
	orf58ng-1	: : : : : : : : : : : EEAETEEAAEAEAAAADTEDIATAVIDNRRIIPFDRSIAECLMPSSESEISFSPVRVPKEITL	130	140	150	160	170
							180
20	orf58-1.pep	EEATRALMSAALRETICKRYIDAPEKNETAVPKVVRVSDTPMEGLQIIGLDDPVVLQRTYSHM					
	orf58ng-1	: : : : : : : : : : EEATRALMSAALRETICKRYIDAPEKNETAVPKVVRVSDTPMEGLQIIGLDDPVVLQRTYSHM	190	200	210	220	230
							240
25	orf58-1.pep	FDAKEAFSEASADYGFEPYFEKQHPSAFSAVKAENARNAPFRRHAGQGKQAEAKSPDVS					
	orf58ng-1	: : : : : : : : : FDAKEAFSEASADYGFEPYFEKQHPSAFSAVKAENARNAPFRRHAGQGKQAEAKSPDVS	250	260	270	280	290
							300
30	orf58-1.pep	OGQSVSQDTAVRNDARRVSVNLKEPNKATVSAEARISRLPESQTVVKRDVEMPSSETEN					
	orf58ng-1	: : : : : : : : OGQSVSQDTAVRNDARRVSVNLKEPNKATVSAEARISRLPESQTVVKRDVEMPSSETEN	310	320	330	340	350
							360
35	orf58-1.pep	OGQSVSQDTAVRNDARRVSVNLKEPNKATVSAEARISRLPESQTVVKRDVEMPSSETEN					
	orf58ng-1	: : : : : : : OGQSVSQDTAVRNDARRVSVNLKEPNKATVSAEARISRLPESQTVVKRDVEMPSSETEN	310	320	330	340	350
							360
40	orf58-1.pep	VFTETVSSVGYPVPPVDETADIHIEPPAPDANVVVEPFEPVPKVPMTAIDIQPFPVSEIY					
	orf58ng-1	: : : : : : : VFTETVSSVGYPVPPVDETADIHIEPPAPDANVVVEPFEPVAVPEDIIDLPPPFPVSEIY	370	380	390	400	410
							420
45	orf58-1.pep	NRTYEPSCFEGVQRSRIAETDHADLVNGWQSEETAAIADDGSEGAAERSSCQYLSET					
	orf58ng-1	: : : : : : NRTYEPSCFEGVQRSRIAETDHADLVNGWQSEETAAIADDGSEGAAERSSCQYLSET	430	440	450	460	470
							480
50	orf58-1.pep	EAFGHDSQAQCPFENVPSERPSCRVSDTDEADCGAFPSSETCAVSHELTDTLLPPLNP					
	orf58ng-1	: : : : : EAFGHDSQAQCPFENVPSERPSCRVSDTDEADCGAFPSSETCAVSHELTDTLLPPLNP	490	500	510	520	530
							540
55	orf58-1.pep	EATQTEEELLENSTITIEKLAEFKVVKVVDSYSGCPVITRYEIEPDVGVRGNSVLNLEKD					
	orf58ng-1	: : : : EATQTEEELLENSTITIEKLAEFKVVKVVDSYSGCPVITRYEIEPDVGVRGNSVLNLEKD	550	560	570	580	590
							600
60	orf58-1.pep	IARSLGCVASIRVVTETPGKTCMGLELPNPKRQMIRLSEIFNSPEFAESKSKTLALCQDI					
	orf58ng-1	: : : IARSLGCVASIRVVTETPGKTCMGLELPNPKRQMIRLSEIFNSPEFAESKSKTLALCQDI	610	620	630	640	650
							660
65	orf58-1.pep	TGQPVVTDLGKAPHILLVAGTTGSGKSVGVNAMILSMLFKAAPEDVRMIMIDPKMLELSIY					
	orf58ng-1	: : TGQPVVTDLGKAPHILLVAGTTGSGKSVGVNAMILSMLFKAAPEDVRMIMIDPKMLELSIY	670	680	690	700	710
							720
70	orf58-1.pep						
	orf58ng-1						

		730	740	750	760	770	780
	orf58-1.pep	EGIPHLLAPVVTDKMLAANALNWCVNEMEKRYRILMSFMGVRNLAGFNQKIAAAARGEKI					
5	orf58ng-1	EGITHLLAPVVTDKMLAANALNWCVNEMEKRYRILMSFMGVRNLAGFNQKIAAAARGEKI					
		730	740	750	760	770	780
		790	800	810	820	830	840
10	orf58-1.pep	GNPFSLTPDDPEPLEKLPFIVVVVDEFADLMMTAGKKIESLILARLAQKARAAGIHLLAT					
	orf58ng-1	GNPFSLTPDDPEPLEKLPFIVVVVDEFADLMMTAGKKIESLILARLAQKARAAGIHLLAT					
		790	800	810	820	830	840
		850	860	870	880	890	900
15	orf58-1.pep	QRPSVDVITGLIKANIPITRIAFOVSSKIDSRTILDQMGAEANLLGGGDMFLLEPGTAYPQR					
	orf58ng-1	QRPSVDVITGLIKANIPITRIAFOVSSKIDSRTILDQMGAEANLLGGGDMFLLEPGTAYPQR					
		850	860	870	880	890	900
20	orf58-1.pep	VHGAFAFASDEEVHRVVBVYLUQFGEPDVDDILSGGGSEELPGIGRSGGDDEDPMYDEAVSV					
	orf58ng-1	VHGAFAFASDEEVHRVVBVYLUQFGEPDVDDILSGGGSEELPGIGRSGGDDEDPMYDEAVSV					
		910	920	930	940	950	960
25	orf58-1.pep	VLKTRKASIISGVORALRIGYNAARALIDQMEAEAGIVSAPEHNGNRNTILVPLDNAX					
	orf58ng-1	VLKTRKASIISGVORALRIGYNAARALIDQMEAEAGIVSAPEHNGNRNTILVPLDNAX					
		970	980	990	1000	1010	

30 Furthermore, ORF58ng-1 shows significant homology to the *E.coli* protein FtsK:

sp|P46889|FTSK_ECOLI CELL DIVISION PROTEIN FTSK >gi|1651412|gnl|PID|d1015290 (D1 division protein FtsK [Escherichia coli] >gi|1651418|gnl|PID|d1015296 (D9072) Cell division protein FtsK [Escherichia coli]) >gi|1787117 (AE000191) cell division protein FtsK [Escherichia coli] Length = 1329
 Score = 576 bits (1469), Expect = e-163
 Identities = 301/459 (65%), Positives = 333/459 (76%), Gaps = 5/459 (1%)
 Query: 556 IEEKLAFKVKVVKVDSYSQGPVITRYEPEIPDVGVRGNNSVLNEKDLARSLGVASIRVVT 615
 +E+ +LA+F+K VV+ GPVITR+E+ GV+ + NL +DLARSL ++RVV
 40 Sbjct: 868 VEARLADFRIKADVVNVYSPGPVITRPELNLAPGVKAARISNLSDLARSLSTVAVRVEV 927
 Query: 616 IPGKTCMGLELPNPKROMIRLSEIFNPSEPAESKSKSLTLAIGQDITPGQVVTDLGKAPHL 675
 IPGK +GLELPN KRL + L E+ ++ #F ++ S LT+ LG+DI G+PVV DL K PHL
 Sbjct: 928 IPGKPVYRNLNNKKRPTVYLRLEVLDNAKFRDNFSPSTLVLGKDIAGEPVVADLAKMPH 987
 45 Query: 676 LVAGTTGSGKGSVGNAMILSMLFKAAPEPDVRMIMIDPKMELLSIYBEGITHLLAPVVTDMK 735
 LVAGTTGSGKGSVGNAMILSML+KA PEDEV IMIDPKMELLS+YEGI HLL VVTDMK
 Sbjct: 988 LVAGTTGSGKGSVGNAMILSMLYKAQPEPDVRMIMIDPKMELLSVYEGIPHLLETVVTDMK 1047
 50 Query: 736 LAANALNWCVNEMEKRYRILMSFMGVRNLAGFNQKIAAAARGEKIGKNNPFFSLTPDDEP-- 793
 AANAL WCVNEME+RY+LMS +GVRNLAG+N+KIAEA I +P+ D +
 Sbjct: 1048 DAANALRWCVNEMERYKLMSALGVNLAGNYKEKIAEADRMRMRFIDDPYWKPGDSMDAQH 1107
 55 Query: 794 --LEKLFPIVVVVDEFADLMMTAGKKIESLILARLAQKARAAGIHLILATQRPSPVWITGL 851
 L-K P+IVV+VDEFADEMLT GKX+SEZIARLAQKARAAGIHL-LATQRPSPVWITGL
 Sbjct: 1108 PVLKKEPVYIWLWDEFADEMLTGVGKKEZELARLAQKARAAGIHLVLATQRPSPVWITGL 1167
 Query: 852 IKANIPTRIAFOVSSKIDSRTILDQMGAEANLLGGGDMFLLEPGTAYPQRVHGAFASDEEV 911
 IKANIPTRIAF VSSKIDSRTILDQ GAE+LLG GDML+ P + P RVHGAF D+EV
 60 Sbjct: 1168 IKANIPTRIAFVSSKIDSRTILDQGAESLLGMGMDLYSGPNSTLPVVRHGAFVRDQEV 1227
 Query: 912 HRVVEYLQKGEPDVOYDILSGGGSEELPGIGRSQGDGETDPMYDEAVSVVLLKTRKASIISG 971
 H VV- K G P YVD I S SE G G G E D+D+AV V + RKASISG
 Sbjct: 1228 HAVVQDWKARGRGPQYVGDITSDSESEGGGG-GFDGAEIDLPLFDQAVQFVTEKRKASIISG 1286
 65 Query: 972 VORALRGINYRAARLIDQMEAEAGIVSAPEHNGNRNTILVLP 1010
 VOR RIGYNAAR+I+QMEA+GIVS HNGRN +L P
 Sbjct: 1287 VORQFRIGYNAARLIIQMEAQGIVSEQGHNGNREVLAP 1325

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 59

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 497>:

```

5      1 ATGATTATTC AAAGAACCT CATCAAAGAA CTCTCTTTA CGCCCGTGG
      51 CATTTCGTC GTCTCTTGG CGGTATTGGT CTCCACCGAG GCAATCAACC
10     101 TGCTCGSCG TGCCTCCGAC GGCG..TGGA TGCCCATCGA TGCGTGTGG
15     151 GCATTGTCG GCTTCTGGT C..... .....
10     901 .....A TTGCCATCGG TTGTTTTTA ATTACCAAA ACGGGCTGAC
951 CCTCTCTTTA GAAGCTCGG AAGACCGAA AATCCATTIT TGCGTGTGGAC
1001 TGCTGCTAT GCACATTATC ATGTTGTCG TTGCACTCAT CCTGTTGGCC
1051 GTCCGAGTA TGCCCAAGCCA GCCCTCTGG CAGGCGTTG CAAAGGTCT
1101 GACATTGAAA GCGGAAATAAAT GA

```

15 This corresponds to the amino acid sequence <SEQ ID 498; ORF101>:

```

1 MIYQRNLKIE LSFTAVGIFV VLLAVLVLSTQ AINLLGRAAD GVIAIDAVL
51 ALVGFWV..... //.
20 301 ...IAIGFLI IYQNGLTLLE EAVEDGKHF WLGLLPMHII MFVLALILLR
351 VRSMPSQPFV QAVGKSILTK GGR*

```

Further work revealed the complete nucleotide sequence <SEQ ID 499>:

```

1 ATGATTATTC AAAGAACCT CATCAAAGAA CTCTCTTTA CGCCCGTGG
51 CATTTCGTC GTCTCTTGG CGGTATTGGT CTCCACCGAG GCAATCAACC
101 TGCTCGSCG TGCCTCCGAC GGCGTGTGG CGATCGATGC CGTGTGTGGCA
151 TTGCTCGGGT CTGCGCTATCG CGGTATGAGC CGUGTTTGCG TGGTGTGAC
201 CGCAATTATC AGTAGCTGAA CGCTGTGAG CGCGTACTGG CGCGCACCG
251 AAATGTCGGGT CTGGCTATCC TGCGGATTTGA CATTGAAACA ATGGATACCG
301 CCGGTGATGA AGTTGGCGT CGCGTTGCG GTTTGGTTC CGCTCATGCA
351 GCTTGTGGT ATACCGGGGG CAGAGCTACG CGAGCGGAA TAGCGCTGAAA
401 TCTTGTGGCA GRAGCAGGAA TTGTTCTTGG TGAGGCGAGG CGAGTTCAC
451 AGTTGGSCA AGCCCAACGG CAGGGTTTAA TTGTCGAAA CCTTCGATAC
501 CGAAATCCGGG ATCATGAAA ACCTGTTCCG GCGGACACAG GACAAAAACG
551 GCGGGCACAC CATCATTCG GCGGAAAGAA GTAATCTTCG CGTGAACGAC
601 AACAAACCGCA CGCTCGAATT CGCCGACCGG TACCGTTACA CGCGCACGCC
651 CGGGACCGCCG GACTACATC AGGTTCCTT CCAAAACACTC AACCTGATTA
701 TCAGCACCATC GCGCAACATC ATGCACCCCCC TTTCACCGG CGTACCATTT
751 CGGACGCCCA ACTGATGGG FAGCAGCAAC CGCGAACACATC AGCGGGAAATT
801 GATGTCGGCGG ATCTGCTGTA CGCTCGACGGT CCTCTACTC TGCTGCTTGG
851 CGCTGGCGCTT TTCTCTGTA ACCGGCGCA CGGGACATAC CTACAAATATC
40     901 TTGATTCGCA TGCGTTCTT TTAAATTAC CAAACCGGC TGACCTGCT
951 TTGTAAGGCC GTGGAAGACG GCAAAATCQA TTTCGCTC GSACTGCTGC
1001 CTATGACAT TATCATGTTT GCGGTGAC CTCATCTGTGTT GCGCGTCCGC
1051 AGTATGCCCC GCGCACCCCT CTGGCGAGCG GTTGGCAAAA GTCTGACATT
1101 GAAAGCGGAA AAATGA

```

45 This corresponds to the amino acid sequence <SEQ ID 500; ORF101-1>:

```

1 MIYQRNLKIE LSFTAVGIFV VLLAVLVLSTQ AINLLGRAAD GVIAIDAVL
51 LVGFIVGIMT PLLLVLTAFI STLTVLTRY RDSEMSWLS CGLAALKQWIR
101 PVMQFAVEFAA VLVAVMOLMV IPWAELRSRQ YAEILKQKQE LSIVEAGEFN
151 SLGKRNRGRVY FVEFTFDTEG IMKNLFLREQ DRNGGDNIIPI AKEGNFSLND
201 NKRLIELRHG YRYSCTPGRA DYNQVSFQKL NLIISTTPKL IDPVSHRTI
251 PTAQALGSSN PQAHELMWR ISITVSVLLL CLLAVPLSYF NPRSGBTYNI
301 LIAIGLFLY QNGTLLFLER VEOKLHLFWL GLPMPHIMF AVAILLLRV
351 SMPSPQFWQA VKSLSLTKGG K*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N. meningitidis* (strain A)

ORF101 shows 91.2% identity over a 57aa overlap and 95.7% identity over a 69aa overlap with an ORF (ORF101a) from strain A of *N. meningitidis*:

		10	20	30	40	50	
5	orf101.pep	MIYQRNLIKEKELSFTAVGIFVVLAVLVLSTQAINLLGRAADGXVIAIDAVIALVGFVWXX					
	orf101a	MIYQRNLIKEKELSFTAVGIFVVLAVLVLSTQAINLLGXAAADXXR-X-AIDAVIALVGFVWXXM					
		10	20	30	40	50	
				//			
10	orf101.pep	90	100	110		
	orf101a	LTVSVLLLCLLAVALPLSYFNPRSGHTYNIKLAIGLFLIYQNGLLFEAVEDGKIHFWGL					
		280	290	300	310	320	330
15	orf101.pep	120 130 140 150					
		LMPMHIMFVIALLILLRVRSPMSQPFWQAVGKSLTIKGGKX					
20	orf101a	LMPMHIMFVIALVLRVRSPMSQPFWQAVGKSLTIKGGKX					
		340	350	360	370		

The complete length ORF101a nucleotide sequence <SEQ ID 501> is:

1	ATGATTTATC	AAAGAAAACCT	CATCAAAGAA	CTCTCTTTA	CCGGCGTCGG		
51	CATTCTTCTGC	GTCTCTCTGG	CGGTATTTGGT	CTCCACCGAC	GCAATCAACC		
101	TGCTCCGCC	TGCCCGCCG	NGGCCTGNTCG	CCATCGATGC	CGTGTGFGCA		
151	TTGCTCGGG	TCTGGTGCNN	NNGNATGACG	CGCGTTTTGC	TNTGTTGTCAC		
201	CGCANTTATC	AGTAGCTTG	CGGTGTTGAG	CGCGTACTGG	CGNGACAGCG		
251	AAATGTCGGT	CTGGGTGATTC	TCCGGATTTGG	CATTGAAACAA	ATGGATACCC		
301	CCGGTGGATG	AGTTTGGCT	CGCGGTGCCC	GTTTTGGTTG	CCSTCATGCA		
351	GCTTTGGGG	ATACCGTGGG	CAGGGTAGCG	CAGCGCGGAA	TACCGTGAAGA		
401	TCCTGAACTA	GAAGCAAGAA	TTGTCCTTGG	TGGAGGAGG	CGGGTTCAAC		
451	AGTTTGGCA	AGCCGACACG	CAGGGTTTAA	TTTGTGCAAAA	CTTCGATAC		
501	CGAATCCTGG	ATCATGAAAA	ACCTGTTTCC	CGCGGACAGAC	CAACAAAAGC		
551	CGCGGACACAA	CATCATTC	CCNAAAGAAA	GTAACTCTTC	GTGAAACGAC		
601	AACAAACCCA	CGCTCGAATT	ATCGCACCGG	TACCGTTTACA	CGCGCACGCC		
651	CGGACGCC	GACTACATC	AGGTTTCCCT	CCNAAAACTC	AACTCTGATTA		
701	TCAGCACCC	GCCCCAACTC	ATCGACCCCG	TTTCCCACCG	CGGTACNATN		
751	CNCACNGCCC	AACTGATTGG	CAGCGACAA	CGCGCACATC	ANCGGGAATT		
801	GATGTGGCG	ATCTCGCTGA	CCGTCAGCGT	CTCTCTACTC	TGCGCTGCTTG		
851	CCGTGGCGCC	TTCCCTATTTC	ACCCCGCCG	CGGACATAC	CTACAAATTC		
901	TTGANTGCCA	TCGGTTGTTG	TTTAATTTCAC	CAAAACGGGC	TGACCTCTGT		
951	TTTGAGAGCC	GTGGAAAGAG	GCRAAAATCCA	TTTTGGCTC	GGACTCTGTC		
1001	CTATGACAT	CATCATGTC	GTCATGSCAA	TGCTACTCTT	GGGGTCCGC		
1051	AGCATGSCCA	GCCAGCCCTT	CTGGCAGGGC	GTGGCAAAA	GTCTGACATT		
1101	GAAGGCGGA	AAATGA					

45 This encodes a protein having amino acid sequence <SEQ ID 502>:

1	MTYQRNLIKE	LSFTAVGIFV	VLLAVLVLSTQ	AINLLGXAA	XRXAIDAVIA		
51	LVGFVWXXMT	PLLVLITAFI	STLTLVTRYW	RDSSEMWSK	CGIAALKQWIR		
101	PVMQFAVFFA	VLVAMQLWV	IPWAEILRSR	YAEILKQKQE	LSLVEAGGF		
151	SLGRKRGRVY	FVEFTFDTEG	IMKNLFLRED	DRNGGDNIIF	XKESNFSLND		
201	NKRTLLELRHG	YRYSCTGPR	DYNQVSFKKL	NLIISTTPRL	IDPVSHRTX		
251	PTAQLGSSS	POXKAEILMR	ISLTIVSLL	CLLAVALPLSYF	NPRSGHTYNI		
301	LXAIGLFLIY	QNLGLLFEA	VEDGKIHFWL	GLLPMHIIMP	VIDAIVLLRVR		
351	SMPSQPFWQA	VGKSLTLKGG	K*				

ORF101a and ORF101-1 show 95.4% identity in 371 aa overlap:

55	orf101a.pep	MIYQRNLIKEKELSFTAVGIFVVLAVLVLSTQAINLLGXAA	DXRXAI	DAVIALVGFVWXXMT	60		
	orf101-1	MIYQRNLIKEKELSFTAVGIFVVLAVLVLSTQAINLLGXAA	DXRXAI	DAVIALVGFVWXXMT	60		
60	orf101a.pep	PLLVLITAFI	STLTLVTRYWRDSEMWSK	CGIAALKQWIR	PVMQFAVFFA	VLFVAMQLWV	120
	orf101-1	PLLVLITAFI	STLTLVTRYWRDSEMWSV	CGIAALKQWIR	PVMQFAVFFA	VLFVAMQLWV	120

	orf101a.pep	IPWAELRSREYAEILKQKQELSLEVAEGGFNSLGRNDRVYFVETFDTESGIMKRNLFRLREQ 180
5	orf101-1	IPWAELRSREYAEILKQKQELSLEVAEGGFNSLGRNDRVYFVETFDTESGIMKRNLFRLREQ 180
	orf101a.pep	DKNGGDNIIFXKEESFLSLDNKRKTLELRHGYRSGTPGRADYNQVSFKLNLIIISTTPKL 240
	orf101-1	DKNGGDNIIFXKEESFLSLDNKRKTLELRHGYRSGTPGRADYNQVSFKLNLIIISTTPKL 240
10	orf101a.pep	IDPVSHRRTXPTAQLIGSSNPQHQAEILMWRISLTSVLLCLLAVALPLSYFNPRSGHTYNI 300
	orf101-1	IDPVSHRRTIPTAQLIGSSNPQHQAEILMWRISLTSVLLCLLAVALPLSYFNPRSGHTYNI 300
15	orf101a.pep	LXAIGLFLIYQNGLTLLFEAVEDGKHFWLGLLPMHIMFVIAIVLLRVRSMPSQFWQA 360
	orf101-1	LXAIGLFLIYQNGLTLLFEAVEDGKHFWLGLLPMHIMFVAVAILLLRVRSMPSQFWQA 360
	orf101a.pep	VGKSITLTKGGK 371
20	orf101-1	VGKSITLTKGGK 371

Homology with a predicted ORF from *N.gonorrhoeae*

ORF101 shows 96.5 % identity in 57aa overlap at the N-terminal domain and 95.1% identity in 61aa overlap at the C-terminal domain, respectively, with a predicted ORF (ORF101ng) from *N. gonorrhoeae*:

	orf101.pep	MIYQRNLIKELSFTAVGIFVVLVLLAVLVTQAINLLGRAADGXVIAIDAVLALVGFWKV 57
	orf101ng	MIYQRNLIKELSFTAVGIFVVLVLLAVLVTQAINLLGRAADGXVIAIDAVLALVGFWVIGM 59
30		//
	orf101.pep	IAIGLFLIYQNGLTLLFEAVEDGKHFWLGL 333
	orf101ng	SLTVSVLLCLLAVALPLSYFNPRSGHTYNIIAIGLFLIYQNGLTLLFEAVEDGKHFWLGL 331
35	orf101.pep	LLPMHIIMFVLA LILLRVRSMPSQFWQAVGKSITLKGKK 373
	orf101ng	LLPMHIIMFVIA I VLLRVRSMPSQFWQAVG 362

The ORF101ng nucleotide sequence <SEQ ID 503> is predicted to encode a protein having partial

40 amino acid sequence <SEQ ID 504>:

1	MIYQRNLIKE LSFTAVGIFV VLLAVLVTQ AINLLGRAAD GRVIAIDAVL A
51	LGWFWIGMT PLLLVITAFI STLTVLTRYW RDSEMSWVLS CGALAKQWIR
101	IPIWAELRSRVE YAEILKQKQE LSLEVAEGEN
151	PVMQFAVFPFA IIIAVMQLWV IPWAELRSRVE YAEILKQKQE LSLEVAEGEN
201	NLKGRNDRVY FVETFDTEGMSV CKMLNFRLREQ DKNGGDNIIF AKEGNFSLKD
251	NKRKTLELRH P YRSTGCGRA ISLTWSVLL CLLAVALPLSYFNPRSGHTYNI
301	STAQLIGSSN PHQAEILMWR ISLTWSVLL CLLAVALPLSYFNPRSGHTYNI
351	LIAIGLFLIY QNGLTLLFEA VELDGKHFWL LLPMHIIMFVIA I VLLRVR SMPSQFWQAVG V...

Further work revealed the complete nucleotide sequence <SEQ ID 505>:

50	1 ATGATTTATC AAAGAACCT CATCAAAGAA CTCTCTTTA CGCCCGTGG
	51 CATTTCGTC GTCCCTCTGG CGGTGTTGGT GTCCACCGAG GGATCAACC
	101 TGTGTCGCC CGCACGCTGCAG GGGCGTGTCG CCATCGATGC CGTGTGGCC
	151 TTAGTCGCC TCTGGCTCAT CGGTATGACCG CGCTTCTTG TGTTGTGAC
	201 CGCACGCTGC CGCACGCTGA CCCTATTGAC CGCGTACTGG CGGCACAGCG
	251 AAATGCGGT CTGGCTCATCC TGCGGATGG CGTTGAAACA GTGGATAGCG
	301 CCCGTCATGC AGTTGGCTG CGCGTTGGCC ATCCCTGATTC CGCTCATGCA
	351 GCTTGTGGTG ATACCGTGGG CAGAGCTCGG CAGCCGGAAATATGCCGAAA
	401 TTTTGAGACCA GAGACGCGAA TCTGTCTTGG TGGAAGCCGG CGAGTTCAAT
	451 AACTTGCGCA AGCGCCACGG CAggtttat TtcgtcgaaA CCTTGGACAC
60	501 CGaactccGC ATCATGAAAA ACCTGTTctc GcgcgAACAG GACAAAAACCG
	551 gccggcacaA CTCATCTTC GCaaaaGAG gtaactTctc gctgaaggac

	601	AACAAA ^{cgc} ca	cgtcgaaTTT	GGGCCACGCC	TACCGTTACA	GCGGcacgcC
	651	CGGacGCGCc	gactaCAATC	AGGTGttcc	cAAAAaAcTc	aacctgATta
5	701	TCAGCACAC	GC ^{CC} AAcATT	ATGaccCGG	TTTCCCACCG	CGGCACCAAT
	751	tgcacGCC	AAcTGATGG	CAGCAGCAAT	CGGCACACATC	AGGCAGAAATT
	801	GATGTCGGCG	ATTCGCTG	CGGTGCGGT	CCTCTGCTC	TGCTCTACTCG
	851	CGGTGCGGT	TTCTATTTC	ACCCGCCGA	GGGACATAC	CTACAAATATC
10	901	TTGATGCGCA	TCGGTTGTT	TTAATTTAC	CAAAACGGG	TGACCCCTGCT
	951	TTTIGAGCC	GTGGAGAGC	GCAAATTC	TTTTGGCTC	GGACTGCTGC
	1001	CTATGACAT	CATCATGTC	GTATGCGAA	TGTTACTCT	GGGGCTCCG
	1051	AGTATGCGCA	GGCAGCCCTT	CTGGCAGGG	TTGGCJAAA	TGCTGACATT
	1101	GAAAGcgGA	AAATGA			

This corresponds to the amino acid sequence <SEQ ID 506; ORF101ng-1>:

	1	MIYQRNL ^L IKE	LSTFTAVGIFV	VLLAVLVSTQ	A1NLLGRAAD	GRVAIDAVL
15	51	LVGF ^W VIGMT	PLLLVLTA ^F I	S1T ^L TVLTRY	RDSEMSVWL	SGLAKQWIR
	101	FVFMQFAVFFA	III ^L AVMQLWV	IPWAE ^L RSRKE	YAEILKQKQE	LSLVEAGEEN
	151	NLGRNRGRVY	FVETFDTESG	IMKNLFLREQ	DKNGGDNIIF	AKEGNFSLKD
	201	NKRTLELRHG	YRYSGT ^F GR	DYNQVSFQKL	NLI ^I STTPKL	IDFVSHRRTI
	251	STAQLG ^I NN	POHQAELMWR	ISLT ^V SVLL	CLLA ^V PLSYF	NFRSGHTYNI
	301	LIAIGLFLIY	QNGLT ^L FEA	VEDGKHF ^W W	GLLPFHIIIMP	VIAIV ^L LRVR
20	351	SMPSFQFWQA	VGKS ^L TLKGG	K*		

ORF101ng-1 and ORF101-1 show 97.6% identity in 371 aa overlap:

	10	20	30	40	50	60
	orf101-1.pep	MIYQRNL ^L IKE	STFTAVGIFV	VLLAVLVSTQ	A1NLLGRAAD	GRVAIDAVL
25	orf101ng-1	MIYQRNL ^L IKE	STFTAVGIFV	VLLAVLVSTQ	A1NLLGRAAD	GRVAIDAVL
		10	20	30	40	50
		70	80	90	100	110
30	orf101-1.pep	PLLLVLTA ^F I	STL ^T TVLTRY	WRDSEMSVWL	SCLG ^L AKQWIR	VFMQFAVFLAVMQLWV
	orf101ng-1	PLLLVLTA ^F I	STL ^T TVLTRY	WRDSEMSVWL	SCLG ^L AKQWIR	VFMQFAVFLAVMQLWV
		70	80	90	100	110
35	orf101-1.pep	130	140	150	160	170
		IPWAE ^L RSRKE	YAEILKQ ^L QKQELS	VLVEAGE ^F NSL ^G KRNGRVY	FVETFDTESGIMKNLFLREQ	
	orf101ng-1	130	140	150	160	170
		IPWAE ^L RSRKE	YAEILKQ ^L QKQELS	VLVEAGE ^F NNL ^G KRNGRVY	FVETFDTESGIMKNLFLREQ	
40	orf101-1.pep	190	200	210	220	230
		DKNGGDN ^I II	FAKEGNFSLN ^D NKTR ^I LEL ^H GYRSGT ^F GRADY	NQVSFQKL	NLI ^I STTPKL	
	orf101ng-1	190	200	210	220	230
		DKNGGDN ^I II	FAKEGNFSLN ^D NKTR ^I LEL ^H GYRSGT ^F GRADY	NQVSFQKL	NLI ^I STTPKL	
45	orf101-1.pep	250	260	270	280	290
		IDPVSHSR ^I ITPA	Q ^L IGGSN ^P QHQAEMLW ^R ISLT ^V SVLL	CLLA ^V PLSYF	NFRSGHTYNI	
	orf101ng-1	250	260	270	280	290
		IDPVSHSR ^I ITPA	Q ^L IGGSN ^P QHQAEMLW ^R ISLT ^V SVLL	CLLA ^V PLSYF	NFRSGHTYNI	
50	orf101-1.pep	310	320	330	340	350
		LIAIGLFLIY	QNGLT ^L FEA	VEDGK ^H FW ^L GLLP	MHIIIMP	VIAIV ^L LRVR
	orf101ng-1	310	320	330	340	350
		LIAIGLFLIY	QNGLT ^L FEA	VEDGK ^H FW ^L GLLP	MHIIIMP	VIAIV ^L LRVR
55	orf101-1.pep	370				
		VGKS ^L TLKGGKX				
	orf101ng-1	370				

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is

predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 60

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 507>:

```

5          1 ..GGTGGTGGTT TTATCAAATGC TTCCGTCGCC ACTTTGACGA CAGCCAAACC
51         GCAATATCAA GCAGGAGACC TTAGCGCTTT TAAGATAAGG CAAGGCAATG
101        TTGTAATCGC CGGACACGGT TTGGATGCGA GTGATACCGA TTACACACGT
151        ATTCTCACTG ATCATTCCAA ATATCGATGCA CCCGTATGG GACAAGATGT
201        TGCTGTCGTC CGCGGACACAA ACCTGATGCC CGCAACAGGT GATGCACATT
251        CGCTTATCTT CAAATAATGCT GCTGCCAATA CGTCAACAAAC TACAGCCAAC
301        AACCGCACAC ATATCCCTT ATTTCGCCATT GATACAGGCA ATTAGGAGG
351        TAT STATGC CAACAAACATC ACCTTGATCA GTACGGTGCAGA GCAAGCAGGC
401        ATTCGCTAA

```

This corresponds to the amino acid sequence <SEQ ID 508; ORF113>:

```

15          1 ..GGGFINASCA TLTTAKPQYQAGDLSAFKIR QGNVVIAGHG LDARDTDYTRILSYHSKIDA 60
51         GGG INAA+ TLT+ P G+L+ F + G VVII G GLD D DYTRILS ++I+A
101        NGTHIPFLFAI DTGKLGGXVC QNQHLDQYGR ASRHS*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with with pspA putative secreted protein of *N.meningitidis* (accession AF030941)

20 ORF and pspA show 44% aa identity in 179aa overlap:

orfl13	GGGFINASCATLTTAKPQYQAGDLSAFKIR QGNVVIAGHG LDARDTDYTRILSYHSKIDA 60
pspa	GGGLINAAVSITLTSVGPVVLNNNLGTFDVSSEKVVIGGKLDTSADYTRILSRAEINA 256
orfl13	PWVGDF/RVVAQGNDVARTGDAHSPILXXXXXXXXXXXXXXGTHIPFLFAIDTGWLGGMYA 120
pspa	VWG-DV-VV-G-N + G + P AIDT LGGMYA 307
orfl13	NKITL1STVEQAGIRNQGWFAASAGNVAVNAEGKLVNTGMIAATGENHAVAHLHARNVHN 179
pspa	+KITL1ST A IRN-G+ FA+ G V ++A+GRL N+G I A +++ A+ V N 362

Homology with a predicted ORF from *N.gonorrhoeae*

ORF113 shows 86.5% identity in 52aa overlap at the N-terminal part and 94.1% identity in 17aa

35 overlap at the C-terminal part with a predicted ORF (ORF113ng) from *N. gonorrhoeae*:

orfl13	GGGFINASCATLTTAKPQYQAGDLSAFKIR 30
orfl13ng	SHPSQLNGYIEVGGRRAEVVIANPAGIVANGGGFINASRATLTTGCPQYQAGDPSGFKIR 224
orfl13	QGNVVIAGHG LDARDTDYTRILSYHSKIDAPWVGQDVRRVAGQNDVAATGDAHSPILNN 90
orfl13ng	QGNAVIAGHG LDARDTDYTRILVCCQNHLDQYGRTSRHS 263
orfl13	IDTGKLGGXVCQQNHLDQYGRASRHS 135
orfl13ng	DFSGFKIRQGNAVIAGHG LDARDTDYTRILVCCQNHLDQYGRTSRHS 263

The complete length ORF113ng nucleotide sequence <SEQ ID 509> is predicted to encode a protein having amino acid sequence <SEQ ID 510>:

```

50          1 MNKTYLRYVIF NRKRGAVVAV AETTKREGKS CADSGSGSVY VKSVSFIPHTH
51 SKAFCFSALG FSLCLALGTW NIAFADGIIT DKAAPKTQQA TILQTGNGIP

```

101 QVNQTPTSA GVSVNQYAF DVGNRGAILE NSRSNTQTQL GGNIQGNPWL
 101 TRGEARVNVN QINSSHFSQL NGYIEVGGR AEVVIANPAG IAVNNGGFIN
 201 ASRATLITGQ PQYQAGDFSG FKIRQGNAVI AGHGLDARDT DFTRILVCQQ
 251 NHLDQYGRTS RHS*

- 5 Based on this analysis, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 61

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 511>:

10 1 ... TCAACGGGAC ATAGCGGARCA AAATTACACT TTGCGCGAG AAATCACACG
 51 CAACATTCTA CTGGCATCAT TTGGCTATGA ATGGCATGTC AAAGCATTAA
 101 GCCATCATGC GCGCACCCAA GCGACTGAGT TGCGCAAAG CAACGGTATT
 151 TGCTTACCTT ATACGTCCAA TTCTTTTACCC CCATTACCCG CGACGACCTT
 201 ATACATTATC ATAATCTGCA ATAAGGCTA TCTTGTGAA ACCGATCCAC
 251 GCTTGGCGA CTACCGCTAA TGGTTGGGTA GTGACTATAT GCGGGACAGC
 301 CTCAACACTG ACCCAACAAAT TTACATAAA CGTTTGGGTG ATGGTTATTA
 351 CGACCAACGT TTATCACTG AACAAATCGC AGAGCTGACAG GGGCATCGTC
 401 GTTTAGAcGG TTATCAAAAC GACGAAGAAATTTAAAGC CTTAATGGAT
 451 AATGGGGCAGC CTGGGGCACG TTtGATGAAAT CTCAAGGGTG GCATITGCATT
 501 AAGTGGCGCA CAAGTGGCGC AACATGACAG CGTATTTGTG TTGTTGGTAC
 551 A AAAAAGAAGT TAAGCTTCTG ATAGGGGGCA CACAAACCGT ATTGGTGC
 601 CAGCGTTTATC TACCCCTTAA AAATGGGAGC ATAGACGCTA AAGGTCGATT
 651 GTTGTCAGGC AGCAATACAC AAATCAATGT TTCAAGGAGC CTGAAAABACT
 701 CAGGCACGAT TGCAGGCGC AACATGCGTTA TTATCAATGCA CGATAACGCTA
 751 GACAAATATCG GTGGCGGTAT TCATGGCGCA AAATCAGCGG TTACGGCCAC
 801 AACAGAACATC AATAATATTG GGGCAAGTGT TTCTGGCGAA CAGACATTAT
 851 TGCTCAACGS AGGCAACACAC ATCACACGCC AAAGCACAC CGCAGCAGT
 901 CAAAATACAC AAGGCAGCAC CACCTACCTA GACCGAATGG CAGGTATTAA
 951 TATCAGACGG AAAAGAAAAC GTGTTT .

This corresponds to the amino acid sequence <SEQ ID 512; ORF115>:

30 1 ... STGHSEQNYTLPREITRNIS LGSFAYEHSRH KALSHHAPSQ GTELPOQSNGI
 51 SLPYTNNSFT PLPESSLLVII NENVKGYLVE TDPRFANVRQ WLGSNDYMLDS
 101 LKLDPNNLHK RLGDGYEQOR LINEQIAELT GHRRLDGYON DEQFKALMD
 151 NGATAARSNN LSVGIALSAE QVAQLTSDIV WLVOKEVKLIP DCGTQTVLVF
 201 QWYVRYRNQD IDKGALLSG SNTQINVSGS LKNSQTIGR NALIINTDTL
 251 DNIGGRHQAQ KSAVATQTDI NNIQGMLSAE QTLLLNAQNN INSQSTTASS
 301 QNTQGSSTYL DRKAQIYITG KEGKV..

Computer analysis of this amino acid sequence gave the following results:

Homology with the pspA putative secreted protein of *N.meningitidis* (accession number AF030941)

ORF115 and pspA protein show 50% aa identity in 325aa overlap:

40 Orf115: 1 STGHSEQNYTLPREITRNIS LGSFAYEHSRH KALSHHAPSQ GTELPOQSNGI 60
 STG+S Y E++ +I +G A+ + + P + + N G I +T
 pspA: 778 STGYSRSPYPAPECV SIRMGISAYKGYAPQQASDIPGTVVPVVAENGIIHPTFT---- 831
 45 Orf115: 61 PLPSSLLVIIINPVNKGYLVEDPRFANVRQWLGSNDYMLDSLKLDPNNLHKRLGDGYEQOR 120
 pspA: 832 -LPNSSLFAIAFPNNKGYLIETDPAFTDYRKWLGSYMLAALQQDPNHIIHKLGDGYEQK 890
 Orf115: 121 LINEQIAELTGHRRLDGYQNDEEQFKALMDNGATAARSMLNSVGIALSAEQAQLTSDIV 180
 50 L+NEQIA+LTG+RRLDGY NDEEQFKALMDNG T A+ + L+ GIALSAEQA+LTS DIV
 pspA: 891 LVNEQIAKLTGYYRLDGYTNDEEQFKALMDNGITIAKELQLTGPIALSAEQAVALRTSDIV 950
 Orf115: 181 WLVQKEVAKLPDPGQTQVLVLPQVIVRVNKGNDIDKGALLSGSNTQINVSGSLKN-SGTAG 239
 WL + V LPDQ TQTVL F+VYVR + D+G+GALLSSG I SG++N G IAG
 pspA: 951 WLENETVTLPDGTTCQVLKPVYVRAPIKDMNGQGALLSGSVDIG-SGAIENRGGIAG 1009

-301-

Orf115: 240 RNALIINTDTLDNIGGRIHAQKSAVTATQDINNIGMLSAEQTLLLNAGXXXXXXXXX 299
 R ALI+N + N+ G + A DI N G + AE LLL A
 pspA: 1010 REALILNAQNPKNIQGDLQKNIFAAAGSDINTTGS-IGAENALLKASNNIESERSETRS 1068

5 Orf115: 300 XXXXXXXXXXXYLDRMAGIYTGKEKG 324
 + R+AGIY+TG++ G
 pspA: 1069 NQNEQGSVRNIGRVAGIYL/TGRQNG 1093

Homology with a predicted ORF from *N.gonorrhoeae*

10 ORF115 shows 91.9% identity over a 334aa overlap with a predicted ORF (ORF115ng) from
N.gonorrhoeae:

15	orf115.pep	STGHSEQNNTLPLREITRNISLGSFAYEHSRK	31
	orf115ng	NEQTFGEKKVFSENGKLHNYWRARRKGHDETGHREQNYNTLPLPEITRDISLGSFAYEHSRK	71
	orf115.pep	ALSHHAPSQGTTELPOSN-----G1SPLPYTSNSITPLPSSSLYIINPVNKGYLVET	81
	orf115ng	ALSRHAPSQGTTELPSQNDRDNIRTAKSNGIISLPYPTNSFTPLPGSSSLYIINPVNKGYLVET	131
20	orf115.pep	DPRFAMYRQWLGSNDMLSLKLDPNNLHKRLGGYGYEQRLINEQIAELTGHRRRLDGYQND	141
	orf115ng	DPRFAMYRQWLGSNDMLSLKLDPNNLHKRLGGYGYEQRLINEQIAELTGHRRRLDGYQND	191
25	orf115.pep	EEQFKALMDNGATAARSMLSVGIALSAEQAQLTS DIVLVLVKEVKLPQGGTQTVLVPQ	201
	orf115ng	EEQFKALMDNGATAARSMLSVGIALSAEQAQLTS DIVLVLVKEVKLPQGGTQTVLMPQ	251
	orf115.pep	VVVRVKNGDIDGKALLSGSNTQINVSGSLKLNCGTIAGRNALIINTDTLDNIGGRIHAQK	261
30	orf115ng	VVVRVKNGGIDGKALLSGSNTQINVSGSLKLNCGTIAGRNALIINTDTLDNIGGRIHAQK	311
	orf115.pep	SAVTAQDINNIGMLSAEQTLLNAGNNINSQSTTAISQNTQGSSTYLDRMAGIYTGK	321
	orf115ng	SAVTAQDINNIGMLSAEQTLLNAGNNINSQSTAKSSQNSQNTQGSSTYLDRMAGIYTGK	371
35	orf115.pep	EKGV	325
	orf115ng	EKGVLAAQAGKDINIIAQGISNQSDQQGQTRLQAGRDIINLTVQTGKYQEIHFDADNHTR	431

An ORF115ng nucleotide sequence <SEQ ID 513> was predicted to encode a protein having amino acid sequence <SEQ ID 514>:

```

1   MLVQTEKDGL HNEQTFGEEK VFSENGKLHN YWRARRKGHD ETGHREQNYT
51  LEETRTRIDS LGSFAYEHSRK KALSRHAPSQ GTELPOSNRD NIRTAKSNGI
101 SLYPTPNST PLPGSSLYII NFANKGYLV E TDFRFAMYRQ WLGSNDMLSL
151 LKLDPNNLHK RLGGYGYEQR LINEQIAELT GHRRRLDGYQ DDEQFKALMD
201 NGATAARSMN LSVGIALSQA QAAQLTS DIVLVLVKEVKLP DGDTQTVLMP
251 QAGNHVRIQT TGTQSSETY HTQTKGSLMS AGIGFTIGSK TNTQENQSQS
301 DNIGGRRIHAQ KSAVATDQD NNIGGILSQA QTLLNAGNN INNQSTAKSS
351 QNAQGSSTYL DRMAGIYTKE KGKVLALA QGKDNINIAGQ ISNQSDGQQT
401 RLQAGDRINL DTVTGQVYI IHFDADNHTI RGSTNEVGSS IQTGKDVDTLL
451 SGNNLNAAKA EVGSSAKGTLA VYAKNDITIS SGTHAGQVDD ASKHTGGRSGG
501 GNKLWITDKA LSHBETAOSS TFEKGQVVLQ AGNDANILGS NVISDNGTRI
551 QAGNHVRIQT TGTQSSETY HTQTKGSLMS AGIGFTIGSK TNTQENQSQS
601 NEHTGSTVGS LKHDITIVS KHYEQTGSNV SSFEGNNLIS TOSMDIGAAQ
651 NQLNSKITQT YEQKGLTVAF SSPVTDLAQQ AIAVAKKAQ QFDKAKTTA
701 MWFLRIPMVG RLFKQAKAPK K*

```

Further work revealed the following partial gonococcal DNA sequence <SEQ ID 515>:

```

1   TTGCTGTGTC AACACGAAA AGACCGTTG CATAACGAGC AAACCTTGG
51  CGAGAGAAAA GTCTTCAGCG AAAATGTTAA GTTGCACAC TACTGGCTG
101 CGCCGCTGAA AGGACATGAT GAAACAGGC ATCTGTAACA AAATTATACT
151 TTGGCAGGAGG AAATCACAC CGAACATTCA CTGGGTCTAT TTGCTCTATGA
201 ATCGCATAGC AAAGCATTAA GCGCTCATGC GCCCAGCCAA GGCACCTGAGT
251 TGCCACAACAA TAACCGGGAT AATATCGTA CTGCGAAAG CAACCGGTATT

```

-302-

301 TCGCTACCTT ATAGCCGCA TTCTTTAACC CCATTCAGGC CGAACGAGCT
 341 ATATCATTAC TAATCTGCCA ATAAGGCTCA TCTTGTGAA AGCGATTCAG
 401 GCTTGGCCAA CCTACCTGGCA TTGGTTGGGA TGACTATATG CTGGCCATTAG
 451 CTCAACATG ACCCAAACAA TTACATATAA CTTGGGGTG ATGGTGTGATT
 501 CGGACGAGT TTATCATGTS ACAAAATCGC AGAGCTGACA GGGCATCTGC
 551 GTTAGCGGG TTATCAACG GAGGAAAGAC ATTAAAGGC CTATGGTGG
 601 AATGGCCCGA CTGGGGAGC TTGATGAAA CTACGGCTG GCATGCGATT
 651 AAGTCGCGG CRAGCAGGC AACCTGACAG CGATATGTT GTGGTGTGAA
 701 AAAAAGAGAT TAAATCTCTT GATGGCGCGA CAAACAGCTT ATGATGCGAA
 751 CAGGTGATAG TACGGCTTAA AAATGGGGCC ATAGACGCTA AGGATGCTC
 801 GTTCTGAGGC AGCAATACAC AAATCACTG TTACAGCAGC CTGAAAACCT
 851 CGGACGAGAT GGGCAGGGCC ATGGCTTAA TTACATGAACT CATTACGCTA
 901 GCAATCATG CTGGGGCTAT TTACGCGCA AATACAGGG CTACGGCC
 951 ACAAAGCATC ATAATATAG GGGCCATTCT TTCTGGCGAG CAGCACATCT
 1001 TCTCTACATG GGTAAACAC CTAACACACC AATACAGCGC CRAGACGAGT
 1051 CAAATACTCAG AAGGTGACAG CACCTTACCTA GACCGAAATGG CAGCTTATTA
 1101 TATCACCGAC AAAAAGAAAAAG TTGTTTGTAGC AGGGCAGCGC GCGAACAGACA
 1151 TCAACATCTG CGGCCGCTTAC ATCAGCAACTC ATACAGATCA GATACGGCTAC
 1201 CGGCTGAGG ATGGGGGGGG CTTAAACATC GATACGGCTAC AGGGGAGCCC
 1251 ATPATCAAGAA ATCCATTAGA TTGGCCGATAA CTCATACATC GATACGGCTAC
 1301 CGAACAGCTT GGGCAGCAGC ATTTCAACAA ACAGGCGATGT TACCCtctGT
 1351 TCAGGGAAATA ATCTCAATG CAAAGCTGCC GAAAGTGGCA AGGGGAAAGG
 1401 CRACCTTGGC GTTGTGCTA AAAATGACAT TATCATCAGC TCAAGCCAAAGG
 1451 ATGGCGGCCA ATGGTGTAGT GGTCACAAAC ATACAGCGAG ACAGCGGGCC
 1501 GGTAAATAT TAGTCATTCG CTGATTAAGG CAAAGTCATC GAAAGTCATC
 1551 TCAAGCAGC ACCTTTGAGA CGAACAGCT TGTTAGTCG CGAGGAAACAG
 1601 ATGGCCAACT CCTTGGCATG ATGTTGTTTT CGGATATGG CACCGGGATT
 1651 CAAGCAGGCA ATCTATGTTG CATGGTGGAC CAAACAACTC AAACGCAAAG
 1701 CGAAAACCTAT CATCAACAGG AAAAATCAGG ATGTTAGTGT CGGAGTGTG
 1751 GCTTCTACATG TGCGACAGC ACAAACACAG AACAAACACAG ATCCCAAAGC
 1801 AACGAACATCA CAGGGCAGTAC CGTAGGCAGC CTGAAAGGGC ATACACCATC
 1851 TTGTCAGCAGC AACAAACTACG CGAACAGCTG CAGCAACCTG TCCAGCCCTG
 1901 AGGGCAACAA CCTTATCGAC CGACGAAAGTA TTGGATATGG CGCCACACAA
 1951 AACAACTTA AGCAGAACAC CACCCAAACCC TGGACAAACAA AGGCTTAAAC
 2001 GTGGCCATTCG AGTTCGGCCS TTACCGATTTG GGGCAGGAAAC GAGGCTTGGC
 2051 TAGCACACA AGCAGAACAC AGAAGCAGACA AACGAAACAAAG GACCCGGCTTA
 2101 ATGGCCATGGC GGCTGACAAAT GCAGGGTTGC AGGGCTTATCA AACAGGCAA
 2151 GGGCCACAAAT ACTTAG

This corresponds to the amino acid sequence <SEQ ID 516; ORF115ng-1>:

40	1 LLVQTEKDGL HNEQTFGEKK VFSENGKLHN YWRARRKGHD ETGHREQNYT 51 LPEEPRDRIS LGGSFAYES KALSRHAPSQ GTELPOSNRD NIRTAKNGL 101 SPLYTFNSTP PLPGSSLYI NPNKANGYLVLE TDPRFANRYQ WLGDYNGL 151 LKLDDNPNLHE RLGDQGYEE LINEQIAELT GHRLRDGYQN DFGCFKALMD 201 NGATAARASLWQ LSVGIALSAL QAAQLTSIDV LWSVKEVPLK DGCTQTVLMP 251 QVYVTRVRXNGO IDGDKRGAQ SNTQINVSQS LKNSTGIAG INTALINTD 301 DNNIGIRRAQO KSAVATQDI NNIGILSSE OTLLLNAGNA INNOSTAKSS 351 QNAQGSTSYL DRMAGIYITG KEGKVLAQQA GKDINIAQG ISNOSDGDT 401 RLOQAGRUDINL DVTCGTQYKE IFHDADNHITI RGTNSVEGGS ICRTGDVILL 451 SGNNILNAKAK EVGSAQGTLI VYAKNNDITIS SGIGHAQVDS AKHTGRSGRS 501 GGNKLWITDKA QSHNEHTTAQS FTEGKQVVLQ AGNDANILIS NVISDNGTRI 551 QAGNHRVLRGT TOTQSOSQTYE HOTOQSGLMS AGIGFTGSS TNFTQENGS 601 NEHTGSTVWS LKGDTTIVAS KHYECTGTSNVN SPEGPNLIS TSQMDIAQG 651 NQNLNSKTTQF YEQGLRGTAVS SSFVPTDLAQOQ AIAVAHKAAKNS KSDKAKTAL 701 MPWRLMLOGVG RFKIQAKAH T*
45	
50	

55 This gonococcal protein (ORF115ng-1) shows 91.9% identity with ORF115 over 334aa:

	20	30	40	50	60	70
	orfl15ng-1.p NEQTFGEKKVFSENGKLNHYWRARRKGHDGETHREQNQYTLPEEEITRDLSGSFAYEHSK					
	::: : : : :					
60						
	orfl15 STGHSEQNQYTLPREITRNSLGSFAYEHSK					
	10 20 30					
	80	90	100	110	120	130
	orfl15ng-1.p ALSRSHAPSQGTLPQSNSRNDRNIRTAKNSGINSLYPTPNSTPLPGSSSLYIINPANKYVLVET					
	: : : : :					
65						
	orfl15 ALSRSHAPSQGTLPQSNSRNDRNIRTAKNSGINSLYPTPNSTPLPGSSSLYIINPANKYVLVET					
	40 50 60 70 80					

	140	150	160	170	180	190
	DPRFANTRYQWLGSYDMLGS	LKLDPPNLLHKRLGDGYB	YBQLRINEQIAELTGHRRLDGYQND			
5	orf115	DPRFANTRYQWLGSYDMLGS	LKLDPPNLLHKRLGDGYB	YBQLRINEQIAELTGHRRLDGYQND		
	90	100	110	120	130	140
	200	210	220	230	240	250
	EQQFKALMDNGATAAARSMNL	LSVGIALSAEQAQLTS	DIWLVQKEVKKLPDGGTQTVLMPQ			
10	orf115	EQQFKALMDNGATAAARSMNL	LSVGIALSAEQAQLTS	DIWLVQKEVKKLPDGGTQTVLVPQ		
	150	160	170	180	190	200
	260	270	280	290	300	310
15	orf115ng-1.p	VYVRVKNGIDGKGALLSGSNT	QINVSGLSKNSGTIAGRNALIINTTD	TLDNIGGRHIAQK		
	orf115	VYVRVKNGIDGKGALLSGSNT	QINVSGLSKNSGTIAGRNALIINTTD	TLDNIGGRHIAQK		
	210	220	230	240	250	260
	320	330	340	350	360	370
20	orf115ng-1.p	SAVTATQDINNIGGILSAEOT	LLLNAGNNINNSQNSTAKSSQNAQGS	STYLDRMAGIYITGK		
	orf115	SAVTATQDINNIGGILSAEOT	LLLNAGNNINNSQNSTTASSQNT	QGSSTYLDRMAGIYITGK		
	270	280	290	300	310	320
	380	390	400	410	420	430
25	orf115ng-1.p	EKGVLAAQAGKDINITIAGQISNQSDQGQTRLQAGRDN	INLDVTQTKYQIEHFADNHTR			
	orf115	EKGV				

In addition, it shows homology with a secreted *N.meningitidis* protein in the database:

30	gi 2623258 (AF030941) putative secreted protein [Neisseria meningitidis] Length = 2273		
	Score = 604 bits (1541), Expect = e-172		
	Identities = 325/678 (47%), Positives = 449/678 (65%), Gaps = 22/678 (3%)		
35	Query: 1 LLVQTEKDGLHNEQTFGEGKVFSENGKLHNNWRAARRKGHDGETGHREQNYTLPEEITRDIS 60		
	+LV T + L N++T G K + ++ G LH Y R +KG D T+G Y E++ I		
	Sbjct: 739 LLVQTEPSALNDETLGTKTID-TDKGLDHRYHRHHKGRDSTGYRSRSPYPEAPEVS- 796		
40	Query: 61 LGSFAYEHSKALSRSRAPSQGTTELQFSNRNIRTAKSNGISLPFYTPNSFTPLPGSSLVII 120		
	+G AY+ + AP Q +++P + + NGI +T LP SSI+ I		
	Sbjct: 797 MGISLAKEYG-----APQQASDIPGTV----FVVVAENGHLFTFT----LNNSSLE 940		
45	Query: 121 NPANKGYLVSTDERFANTRYQWLGSYDMLGS	LKLDPPNLLHKRLGDGYB	YBQLRINEQIAELT 180
	+ P NGKLY+ETDP F +Y+RLGS Y +L+ DEN++HKRLGDGYYEQ+L+NECIA+LT		
	Sbjct: 841 APNNKGYLIETDPAFTDYRKWLGSYMLAQLQDPNHNHKRLGDGYYEQKLVNNEQIAKLT 900		
50	Query: 181 GHRRLDGYQUNDEEFKALMDNGATAAARSML	LSVGIALSAEQAQLTS	DIWLVQKEVKKLP 240
	G+RLDGY NDEEFKALMDNG T A+ + I+ GIALSAQG A+LTSDIWL + V LE		
	Sbjct: 901 GYRRLDGTYNDEEFQKALMDNGTTIAKELOITGIALSAEVARLTS	DIWLVLENETVLF 960	
55	Query: 241 DGGTQTVLKVYVVRKNGIDGKGALLSGSNT	QINVSGLSKNSGTIAGRNALIINTTD 299	
	DG TQTVL P+VWV + +G+GALLSGS I SG++N G IAGR ALI+N		
	Sbjct: 961 DGTTQTVLKVYVVRKMDNGQQGALLSGSVV+DG-SGAIENRQGGLIAGREALIINAQN 1019		
60	Query: 300 LDNINGGRHIAQKSAVTATQDINNIGGILSAEOT	LLLNAGNNINNSQNSTAKSSQNAQGSSTY 359	
	+ N+ G + + A DI N G I AE LLL A NNI +S +S+ON QGS		
	Sbjct: 1020 IKNLQGDLQGKNIFAAAGSDITDTNGTSI-GAENALLLKASNTIESRS	SETTRSNQNQECSVNR 1078	
65	Query: 360 LDRMAGIYITGKEKGVLAAQAGKDINITIAGQISNQSDQGQTRLQAGRDN	INLDVTQTKYQ 419	
	+R+AGIY+TG++ G + AG + I + A +++NGS+ GOT L AG DI DT + Q		
	Sbjct: 1079 IGRVAGIYLTRQNGSVLLDAGNNIVITASELTNQSEDGQTVLNAGGDIRS	DTTGTISRNQ 1138	
70	Query: 420 EIIFHDADNHTRGSTGNSNEVGSSGTCGDTLISGNLNAKAA	SVGSAGKTLLAVYAKNIDITI 479	
	F+D+IR NEVGS+I-T+G++ + + +AEVGS +G I + A DI +		
	Sbjct: 1139 NTIFDSNDYVIRKEQNEVGSTIRTGRNLNLSNAKGDIRIRAAEVGSEQRKLLAGRDKIV 1198		
	Query: 480 SSGIHAQGVDDASKHTGRSGGGNKLVLITDKAQS	SHHETAQSSSTFEGKQVVIQAGNDANILG 539	
	+G + +DA K+TGRSGGG K +T ++ A S T +GK++L +G D + G		
	Sbjct: 1199 EAGKAHETEDEDALKYTGRSGGGKMRHLLKRNQNQGAVSGTLDGKEIIILVSGRDITVTG 1258		

Query: 540 SNVISDNGTRIQAQHNVRIGTTQTSQSETYHQTQKSGLM-SAGIGFTIGSKTNQENQS 598
 SN-1+DN T + A N++ + T+S+ + +RSGLM S GIFT GSK +TQ N+S
 Sbjct: 1259 SNIIDAHNTILSAKNNIVLKAETRSRSAEMNKKERSGLMGSGGIGFTAGSKKDQTNRNS 1318

5 Query: 599 QSNEHTGSTVGSLKGFTTIVASKHYEQTGSNVSSPEGNNNLISTQSMDIGAAQNQLNSKT 658
 ++ HT S VGSL GT+ I A KHY CTGS +SSP+G+ IS+ + I AAQN+ + ++
 Sbjct: 1319 ETVSHTESVVGSLNGNTLISAGKHYQTGSTISSLSPQGDVGVISSGKISIDAQNRYSQESK 1378

10 Query: 659 QTYEQKGLTVAFSSPTD 676
 Q YEQKG+TVA S PV +
 Sbjct: 1379 QYEQKGVTVVAISVPVN 1396

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 62

15 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 517>:

```

 1 ..TCAGGGAATA ACCTCAATGCC CAAAGCTGCC GCGCAAACGG
 51 TACACTCGGT GTGTCATGCC ATAATGACAT CAACATCAGC GCAGGCATCA
 101 ACAGACCCAA TGTTGATGAT CGCTGCAAC ACACAGGCA AGGGGTGGT
 151 GCGCAATTAA TAGCTTATTAC CGATAAAGGC CAAAGTCATC ACGAAACCG
 201 CCAAAAGCACG ACCCTTGAAC CGAACGCAAGT TGTTATTCAGC GCAGGAAACG
 251 ATGCGAACAT CCTTGAGCAGC ATATGTTATT CGGATAATGG CACCCAGATT
 301 CAAGCAGCAC ATCATGTTCC CATTTGTTACA ACCCAAACTC AAAGCCAAG
 351 CGAAACACTT CATCAACACCC AGAAATCAGG ATTGTGAGT GCAGGTATCG
 401 GCTTCATAT TGCCAGCAAG ACAAAACACA AGAAAACACCA ATCCCAAAGC
 451 AACGAAACATA CAGGCCAGTAC CGTAGGCAGC TTGAAAGGGC ATACCACAT
 501 TTGTCAGGCC AACRACTACG AACRAANTCG CAGTACCGTT TCCAGCCCCG
 551 AAGGCCAAACA TACCATCTAT GCGCAANCA TAGACATTC AAGCGGCACAC
 601 AACAAATTAA ACAGTAATAC GACCCACCC TATGAAACAA AGGG. CTAAAC
 651 GGTGGCATTC AGTTGGCCCG TTACGAGTT GGCAACACAA ...

```

30 This corresponds to the amino acid sequence <SEQ ID 518; ORF117>:

```

 1 ..SGNNLNAAKA EVSSANGTLA VSANNDINIS AGINTTHVDD ASKHTGRSGG
 51 GNKLVITDKA QSHIIHETQASS TFEKGKVVLQ AGNDANILGS VNISDNGTQI
 101 CAGHNVRIGT TOTQSQSETY LTQCKSGLM S AGIGFTIGSK TNTQENQS
 151 NEHTGSTVGS LKGDTTIVAG KHYEQIGSTT SSPEGNNNTY AQSIDIQAAH
 201 NKLNSNTTQD YEQKQLTVAF SSPVTDLAQ ...

```

Computer analysis of this amino acid sequence gave the following results:

Homology with the pspA putative secreted protein of *N.meningitidis* (accession number AF030941)

ORF117 and pspA protein show 45% aa identity in 224aa overlap:

40	Orf117: 4	NLNKAAEVSSANGTLA VSANNDINIS AGINTTHVDD ASKHTGRSGG
	pspA: 1173	++ +AAEV S G L ++A DI + AG T +DA K+TGRSGGG K +T ++
45	Orf117: 64	HETAQSSSTFEGKQVVLQAGNDANILGSVNISDNGTQIQAQHNVRIGTTTQTSQSETYHQT 123
	pspA: 1233	+ A S T +GK++L +G D + GSN+i+DN T + A N++ + +T+S+ + +
50	Orf117: 124	NGQAVSGLTLDGKEIILVSGRDIITVTGSNIIIAHNHTILSAKNNIVLKAETRSRSAEMNKK 1292
	pspA: 1293	+KSGLM S GIFT GSK +TQ N+S++ HT S VGSL GT+ I AGKHY Q GST+SS
	Orf117: 183	EKSGLMGSGGGIGFTAGSKKDQTNRSETSVSHTEVVGSLNGNTLISAGKHYQTGSTI 1352
	pspA: 1353	PEGNNTIYQAQSIDIAQAHNKLNSNTTCTYEQKQLTVAFSSPTD 226
		+P+G+ I + I I AA N+ + + Q YEQK +TVA S PV +
		PQGDVGIVSSGKISIDAQNRYSQESKQVYEQKGVTVVAISVPVN 1396

Homology with a predicted ORF from *N.gonorrhoeae*

ORF117 shows 90% identity over a 230aa overlap with a predicted ORF (ORF117ng) from *N.gonorrhoeae*:

5	orf117.pep	SGNNLNAKAAEVSSANGTLAVSANNNDINIS	30
	orf117ng	HFDADNHHTIRGSTNEVSSIQTKGDVTLSSGNNLNAKAAEVGSAKTLAVYAKNDITIS	480
10	orf117.pep	AGINTTHVDDASKHTGRSGGGNKLVITDKAQS HHEHTAQSSTEPGKQVVLQAGNDANILGS	90
	orf117ng	SGIHAMGQVVDASKHTGRSGGGNKLVITDKAQS HHEHTAQSSTEPGKQVVLQAGNDANILGS	540
15	orf117.pep	NV1SDNGTRI1QAGNHRVIRGTTQ7QS5QSETYHQTQKSGLMSAGIGFTIGSKTNTQENQSQS	150
	orf117ng	NV1SDNGTRI1QAGNHRVIRGTTQ7QS5QSETYHQTQKSGLMSAGIGFTIGSKTNTQENQSQS	600
20	orf117.pep	NEHTGSTVGSGLKGDTTIVASKHYE2Q1GSTVSPEGNNTIYAQSIDIQAHHNKINSNTQT	210
	orf117ng	NEHTGSTVGSGLKGDTTIVASKHYE2Q1GSTVSPEGNNTIYAQSIDIQAHHNKINSNTQT	660
	orf117.pep	YEQKXLTVAFSPVTDLAQQ	230
	orf117ng	YEQKGLTVAFSPVTDLAQQAIAVAHHKAQFDKAKTTALMPWRLPMQVGRLFQKAKPK	720

An ORF117ng nucleotide sequence <SEQ ID 519> was predicted to encode a protein having amino acid sequence <SEQ ID 520>:

25	1	.LLVQTEKDGL HNEQTFGEKK VFSENGKLHN YWRARRKHD ETGHREQNYY	
	51	LPEEITRDIS LGSFAYEHS KALSRHAPSQ GTELFQSNRD NIRTAKSNCI	
	101	SLPYTPNNSFT PLPGSSLYII NPANKGYLVE TDPRFANYRQ WLGSDDYMLGS	
	151	LKLDPNNLHK RLGDGYLEQR LINEQIAELT GHRRLDGYQN DEEQFKALMD	
	201	NGATAARNSM LSVGIALSSE QAAQTLSDIV WLVRQEVKLLE DGGTQVILMP	
30	251	QVYVVRVKNGD IDPGKALLSG SNTQINVSSG LKNNSTIAGR NLLINNTDTL	
	301	DN1IGRGRHQAQ KSAVLLTQD1 HNNIGLSSA QTLLNNAGQ INNOSTAKSS	
	351	QNAGQGSSYT DRMAGIYTQG KEGKVLAAQD GKDIINIIAQG ISNOSDQGOT	
	401	RLLQAGRDRIDL DTVQPGKYQZ IHFADNHTI RGSTNEVGSS IQTKGDVFTLL	
35	451	SGNNLNAKAA EVGSAKGTLL YAIAKNDITIS SGHAGQVDD ASKHTGRSGG	
	501	GKNLVITDQA QSHHETATQSS TFEKGKQVVLQ AGNDANILGS NV1SDNGTRI	
	551	QAGNHRVIRGTT QTQ5QSETYHOTQKSGLMS AGIGFTIGSK TNTQENQSQS	
	601	NEHTGSTVGS LKGDTTVAF KHYE2Q1GSTVSPEGNNTIYAQSIDIQAHHNKINSNTQT	
	651	NQLNSTKTTQT YEQKGLTVAFSPVTDLAQQAIAVAHHKAQFDKAKTTALMPWRLPMQVGRLFQKAKPK	
	701	MFWPLRMQVG RLFQKAKPK K*	

40 Further work revealed the following gonococcal partial DNA sequence <SEQ ID 521>:

45	1	TTGCTTGTGC AAACAGAAAA AGACGGTTTG CATAACGAGC AACACCTTTGG	
	51	CGAGCAAGAA GCCTTCACGG AAAATCTGTA TTGTCGACAC TACTGGCGTG	
	101	CGCGCTGTTAAG AAGCACGAT GAAACAGCGG ATCTGTGACA AAATTATACT	
	151	TTGCGCATAGC AAATCCTAACG GAACATTTCGA CTGGGTTCCAT TTGCGCTATGTA	
50	201	ATCGCCATAGC AAAGACATTA GCGCTCATGC CGGCCAGGCCA GGCACTGAGT	
	251	TGCGCAAAAGG TAACCGGGGAT AATATCCTGTA CTGCGAAGAAC CACCGTTATT	
	301	TCGCTTACCTT ATACGCCCAA TTCTTCTTAC CCAATTACCC CGACGAGCTT	
	351	ATACATTTAC ATATCTGCCA ATAAGGCTA CCTTGTGTAAC ACCGATCCAC	
	401	GCTTGTGCCA CTACCGTCAA TGTTGTGTTA GTGACTATAT GCTCGCCAGC	
	451	CTCAAACTAG ACCAAAACAA TTTCATATAA CGTTTGGGTG ATGGTTATTA	
	501	CGAGCAACGT TTAATCATG ACAAATCTGC AGAGCTGACA GGGCATCGTC	
	551	GTTTGTAGCGG TTATCAAAAG GACGAAAGAAC AATTAAAGC CTTAATGGAT	
	601	ATATGGCGGCA CTGCGGCAAG TTCTGATGAACT CTGACGCTTG SCATTCGATT	
55	651	AAGTGGCGCA CAAGCAGCGC AACTGACCAA CGATATTTGTG TGGTTGTGAC	
	701	AAAAGAAGAT TAACTCTCTGATGCGGCCA CACAAACCGT ATTGATGCCA	
	751	CAGGTTTATA TACGGCTTAA ATATGGCGC ATAGACCGTA AAGGGTCATT	
	801	GTTGTCAGGC AGCAATACAC AAATCAATGT TTTCAGCGACG CTGAAAACACT	
	851	CAGCGCCGAT TGCAGGGCGC ATAGCGCTTA TTATCAATAC CGATACGCTA	
	901	GACAATATGC TTGGCGCTAT TCATGGCCAA AAATCAGCGG TTACGGCCAC	
60	951	ACAGACATC ATAATATAA TGCGCCATCTT TTCTGCGCAA CAGACATTAT	
	1001	TGCTCAATGC GGGTAACAC ATCAACAAAC AAAGCACGGC CAAGACGACT	
	1051	CAAAATGCAC AGGTAGCGC CACCTACCTA GACCGAATGG CAGGTATTTA	

1101 TATCACAGGC AAAGAAAGAG GTGTTTTAGC AGCGCAGGCC GCGAAAGACA
 1151 TCAACATCAT TGGCGTCGAAT CAGCAATC ATCAGATCA AGGGCAACC
 1201 CGGTGCAAGG CAGGACCGGA CATTAACTTG GATACGGTAC AAACCGGCA
 1251 ATATCAGAA A'CCATTGATG ATGCCGATATA CCATACCCATC CGAGGTTCAA
 5 1301 CGAACAAAGT CGCGCAGCAG ATTCAAACAA AAGGGATGT CTTCatATG
 1351 TCAGGGAAAT ATCTCAATGC CAAAGCTGCA GAAGTCGGCA GCGAAAAGG
 1401 CACACTGCT GTGTTATGCTA AAATGACAT TCTATTCAGC TCAAGGCATCC
 1451 ATGCCGCGCA AGTTGATGAG CGTGTCAAAC ATACAGGGCAG AAGCGGCGC
 1501 GGTAATAAAT TAGTCATTC CGATAAAAGC CAAAGTCATC AGAAACTGC
 10 1551 TCAAGACAGT ACCTTTGAAG CAAAGCAAGT TGTTATTCAG CGAGGAAACG
 1601 ATGCCAACAT CCTTGGCAAT ATGTTATTG CGGATAATGG CACCGGATT
 1651 CAAGCAGGA ATCATGTTGC CATTTGATACA ACCCAAACTC AAAGCCAAG
 1701 CGAACAACTAT CATCAACAA AAAATCAGG ATTGTAGATG CGAGGTATCG
 1751 GCTCTACTAT TGGCAGCAAG CAAACACACAG AAGAAAACCA ATCCAAAAGC
 1801 AACGAAACATA CAGGCACTAC CTGAGGCAGC CTGAAAGGCC ATACACCAT
 1851 TGTGGAAGA AACAAACATC CAAACACCGG CAGCAACGTT TCCAGGCCCTG
 1901 AGGGCAACAA CCTTATCAGC AGCGAAAGTA TTGATATTCG CGCAGCACAA
 1951 AACCAAATTA ACAGCAACAC CACCAAAAC TAGGRACAAA AAAGCTTAA
 2001 GGTGGATTTC AGTTGGCCCC TTACCGTTT GCACACACAA CGGATTCCCG
 20 2051 TAGCAACACNA AGCAGCAACAG AAGTCCGACAA AGCAAAAMC GACCGGCTTA
 2101 ATGCACTGGC GCGTCACAACT GCGAGGTGGC AGGCCTATCR AACAGGCCAA
 2151 GGGCCAAAAA ACTTAG

This corresponds to the amino acid sequence <SEQ ID 522; ORF117ng-1>:

1 LLVQTEKDGL HNEQTGEKK VFSENGKLHN YWRARRKHD ETGHIREQNYT
 25 51 LPEEITRDIS LGSFAYEKS KRALHRAPSQ GTELPEQSNRD NIRTAKNSGI
 101 SFLPYTNSFTI PLEGPSYLII NPANKYLVLE TDPRFANYR WLGSDDYMLGS
 151 LKLKPBNLHLR RLGDQGYEQR LINEQIAELT GHRRLDGYON DEEQFKALMD
 201 NGATAVARSMA LSVGIALSSE QAAQLTSDIV WLWQEVKLP DGGTQTVLMP
 251 QVYVRRHQAQ IDKGKALLSG SNTCINVSGS LKNSTIAGR HALLINTDTL
 30 301 DNIGRGIHAQ KSAVTAQTDI HNIGLILSE C'LLLNAGNN INNOSTAKSS
 351 QNAQGGSYTL DRMAGIYT KEGKVLRAAQ GKDINIIAGQ ISNOSDQQT
 401 RLQAGRDIINL DITVQGYKOE IHDFAANHII RGSTMNEVGSS IJIKGDGVTL
 451 SGNNIAKAA EWGSQATLVA VVAKNDITIS SGIIHGCGDD ASKHTGRSGG
 501 GNRKVITDQA QSHHETAQSS TFEGKQVLIQ AGNDANILGS NVISDNGTRI
 551 QAGHNRVIGT TQTPQSSETY HQTKSGLMS AGIGFTIGSK TINTQENQSOS
 601 NEHTGSTVSX LKGDTSETVS XHYEQTGSNV SSPEGNNLIS TQSMIDGAAQ
 651 NQLNSKTTTGT YEQQGLTVAF SSFVTDLQQ AIAVAHKAAK KSDRAKTTAL
 701 MWRLPLPMQV RFIQKAHHK *

ORF117ng-1 shows the same 90% identity over a 230aa overlap with ORF117. In addition, it

40 shows homology with a secreted *N.meningitidis* protein in the database:

```
gi|2623258 (AF030941) putative secreted protein [Neisseria meningitidis] Length = 2273
Score = 604 bits (1541), Expect = e-172
Identities = 325/678 (47%), Positives = 449/678 (65%), Gaps = 22/678 (3%)
45 Query: 1 LLVQTEKDGLINEEFTGKVFSENGKLHN YWRARRKHD ETGHIREQNYT LPEEITRDIS 60
  L+N+T G K + ++ G LII R +KG D TG+ Y E++ I
Sbjct: 739 LIVGTFESALONDDETLGKTI-TDKGDILYHHRHHKGRDSTGYRSRSPYEPAPEV-SIR 796
50 Query: 61 LGSFAYEKS KRALHRAPSQGTEPQSNRD NIRTAKNSGILSYTPNSFTPLFEGSLVII 120
  +G AY+ + AP Q +++P + + NGI +T LP SSL+ I
Sbjct: 797 MGISAYKGY-----APQKASDIPGTV-----VPVVAENGIHPTF-----LPNSSLFAI 840
55 Query: 121 NPANKYLVLE TDPFPFANRQNLGSDDYMLGS LKLDPNNLHKKRLGDGYEQR LINEQIAELT 180
  P NKGYL+ETDP F +YR+NLGS YML +L+ DPN++HKRLGDGYEYQ+L+NEQIA+LT
Sbjct: 841 ANNNKGYLIEETDPAFTDYRKWLGSGYMLAALQDPNNIIHKKRLGDGYEQQKLVNNEQIAKT 900
60 Query: 181 GHRRLDGYQNDEEQFKALMDNGATAARSMSLNSVGLIA SAEQAAQLTSDIVWLWQEVKLP 240
  GRRRLDGY NDEEQFKALMDNG T A+ + L+ GIALSAEQ A+LTSIDVNL + V LP
Sbjct: 901 GYRRLDGYNDEEQFKALMDNGITIAKEQLQTGPGLSVAEVARLTSIDIVWLNEVTVLP 960
65 Query: 241 DGGTQTVLMPQVYVRVRNNGIDKGKALLSGNTQINVSGSLKN-SGTIAGRNLAIINTD 299
  DG TQTVL P+YVYR + ++G+GALSGS I SG++N G IAGR ALIN
Sbjct: 961 DGGTQTVLKPVKVYVRARPKDMMNGQGALLSGSVVDIG-AGAIENRGGLIAGREALILNQN 1019
Query: 300 LDNIGGRIHAQKSAVTATQDINNIGGILSAEQT LLLNAGNNINNQSTAKSSQNAQGSSTY 359
```

+ N+ G + + A DI N G I AE LLL A NNI ++S +S+ON QGS
 Sbjct: 1020 INKLQGDLQGKNIIFAAGSDITNTGSI-GAENALLKASNNIESRSETRSNQNEQSVRN 1078
 5 Query: 360 LDRMAGIYITKGKEKGVLAAQAGKDINIIAQGQISNQSDQGOTRLQAGGRDINLDTVQTGKYQ 419
 + R+AGIY+TG++ G + AG + I + A +++NQ+ GOT L AG DI DT + Q
 Sbjct: 1079 IGRVAGIYILTGRQNGSVLLDAGNNIVLTASELTLNQSEDGOTVLNAGGDIRSDDTTGISRNQ 1138
 Query: 420 EIHEFDADNITIRGSTNEVGSSIQTKGDVTLLSGNLNNAKAREEVGSAKGTVAEYAKNDITI 479
 FD+DN+ IR NEVG+S+I++G++L + ++ +AEVGS +G L + A DI +
 10 Sbjct: 1139 NTIFDSNVYIRKEONEVEGSTIRTRGNLNSLAKGDIRIRAAEVGSEOGRLKLAAGRDKIV 1198
 Query: 480 SNSIHAGQVDDASKHTGRSGGGKLVLAQSHETAOSSSTFEGKOVVLLQAGNDANILG 539
 +G + +DA K+TGRSGGG K +T + + + A S T +GK++L +G D + G
 Sbjct: 1199 EAGKAHTETEDDALKYTGRSGGGIKQKMTRHLKQNQAVSGTLDGKEIIIIVLVSGRDITVTC 1258
 15 Query: 540 SNVISDNGTRIQAGNHWRIGHTTQOQSOSETYHOTKSGLM-SAGICFTIGSKNTNOENS 598
 SN+I+DN T + A N++ + + + +S+ + +KSGLM S GIFT GSK +TO N+S
 Sbjct: 1259 SNITADNHITLSSAKNNIVLKAATRBSRAEMNKKEKSGLMGSGGIGTAGSKKDTQYRN 1318
 20 Query: 599 QSNEHTGSTVGSLLKGDTTIVASKHYEOTGSNVNSPFGNNLISTQSMDIGAQNQLNSKTT 658
 ++ HT S VSGL GT I A KHY QTGS +SSP+G+ IS+ + I AAQN+ + +
 Sbjct: 1319 ETVSHTESVGSGLNNTLISAGKHYTQGSTISSLPGDVGISSGKSIDADAQNRYSQESK 1378
 25 Query: 659 OTYQOKGLTVAFSSPTVD 676
 O YEQKG+TVA S PV +
 Sbjct: 1379 QYEQKGVTVAISPVVNN 1396

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 63

30 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 523>:

```

1 ATGATTTACA TCGTACTGTT TCTAGCTGTC GTCCCTGCCG TTGTCGCCCTA
51 CAACATGTTA CAGGAAAACC AATACCGCAAA AAAAGTCGGC GACCAAGTTGG
101 GACACTCCGA CAAAGATGCC CTGCTCAACA GCAWAAACCG CCATGTCGCC
151 GACGGCAAAAC CGTCGGCGGG CGTCAGTCATG ATGCCGAAAC CCCAACCGGC
201 GGTCAAAAAAA CGCCGAAACCC CGCAAGACCC CGYCATGGCC AACCTGCAAG
251 AACAGGATGC CGTCATCACCG GCCAAGCAGA AACAGGCAAA ACCTCTCCCG
301 TTCAAAACCG AACATGGACCG CGCTCTTGGNA GAAAGCCGCA TTATCCGGCA
351 CTCCGGCCAC ACCGTTTCG ACCCCAAAC CGGACATTCG GCAACGAAAC
401 CTGGCGGAGC GTGCCGAAAAA CCTGCACCGG TTCCGCAAAC ACCTGCAAAA
451 CGCTGATTA CGCTCAAGA ACTGTCAAAAGA GTGCAATTAT CCTGGTTTGA
501 CGTGGCGCATC GACTTCATCT CCTT...
  
```

This corresponds to the amino acid sequence <SEQ ID 524; ORF119>:

```

1 MIYIVLFALV VLAVVAVNYM QENQYRKKVR DQFGHSKDAA LLNSXTSHVR
51 DGKFGGGVSM MKFKPFPVAKK TAKPDXPMR NLQEODAVYI AKQKQAKASP
45 FKTETIETALE ESGIIGNSAH TVSEFOTGHIS ATKPADASAK PAPVPQTPAK
101 PLITLKELSK VELSWFDVRI DPISY...
  
```

Further work revealed the complete nucleotide sequence <SEQ ID 525>:

```

1 ATGATTTACA TCGTACTGTT TCTAGCTGTC GTCCCTGCCG TTGTCGCCCTA
51 CAACATGTTA CAGGAAAACC AATACCGCAAA AAAAGTCGGC GACCAAGTTGG
101 GACACTCCGA CAAAGATGCC CTGCTCAACA GCAWAAACCG CCATGTCGCC
151 GACGGCAAAAC CGTCGGCGGG CGTCAGTCATG ATGCCGAAAC CCCAACCGGC
201 GGTCAAAAAAA CGCCGAAACCC CGCAAGACCC CGYCATGGCC AACCTGCAAG
251 AACAGGATGC CGTCATCACCG GCCAAGCAGA AACAGGCAAA ACCTCTCCCG
301 TTCAAAACCG AACATGGACCG CGCTCTTGGNA GAAAGCCGCA TTATCCGGCA
351 CTCCGGCCAC ACCGTTTCG ACCCCAAAC CGGACATTCG GCAACGAAAC
401 CTGGCGGAGC GTGCCGAAAAA CCTGCACCGG TTCCGCAAAC ACCTGCAAAA
451 CGCTGATTA CGCTCAAGA ACTGTCAAAAGA GTGCAATTAT CCTGGTTTGA
501 CGTGGCGCATC GACTTCATCT CCTT...
  
```

651 CTTATCAGCGA TTTATGCTGGG GTAACTCAGGC AGTCAGCGCG AGTCAGACTG
 701 CCTTCCGAGCA AGAACATCTG GCAATTCTGGG CGAACCTGGG CGACCTGGCGA
 751 CAAAGCATGG CGGGTCAGAC GCTTGACACC GACCTTGTGGG CCTTTATCGAA
 801 ATGGTCTTCG CCAACTGAGC CATTCTGGCG CGCGCTGCGAC CAGGACATCG
 851 CCTTCATCTT GGTGGTCCCCG ACCAGCATGC CGGGCTGAGA ATCGCTGGCG
 901 CGCGTAAAGG CGGTGGGGGG CCTGGTTGGAA GAGGACGGGG CGTTCACATA
 951 TACCGAACAGG CGGGCTGCTGA CCTATGTTCTC CACTCTGCTG
 1001 AGCGCCATTAC CAAGCCGGCT TTGGCAACACC AGTCCTACAG AGCTCTCAGT
 1051 ATSTCTGCTC3 ACATCCGGCA CTCTCGCGCA CGGGAAAAAA CCTTCGCGCA
 1101 TTGTGTTTAA GATTGGCGG TGACGGCTGTC CGGCCGATGTC AAGTCGATCT
 1151 TGCTCAACAGA AAAATGGGA GAAAGTTGGCA CAAATGGCT CAAAGAGCTG
 1201 CGCACTTATG TATTGGCCGG TCAGTCGGAG ATGCTCAAAAG TCGGGATTCGA
 1251 ACGGGCGCGC AAAAGGGCGAT CGGGCTGTGT CCTCTTAA

This corresponds to the amino acid sequence <SEQ ID 526; ORF119-1>:

15 1 MIYIYLFLAV LVALVAYNNY QENQYRKVR DQFGHSKDKA LLNSKTSHVR
51 DGKPQGSVSM MPKQPQAVVK TAKPQDPMR NLQEQDQAVYI AKQOKQAKAS
101 FKTTEIATLAE ESGILGNSAH TVESEPQTGHG APKRPADAPL PAVPVQTPKAP
151 PLTILKLSEK VELFWFDWPR DFISVIALTE AKELHALFRPL SRRNCYQV
201 CTMDHDFQIA EPIPGRYQA FIVIGVQASR NLQASLELS AFNQNDVQDFA
251 QNSGGQTHTL DLAIFAEVAS ALDAFCAARVD QTIAHLVTS S7ISVGLERS
301 ATVGVFGLBE DDGAGHYTUD SGSTMFSICS NLENPFNTAUN LUNQSYQGF
351 MLLDOPHSB GKBTFDLMLPA DLAVRLRSQGL NLNLVNNDKME EVSTQWLKD
401 RTVYLRQSE LMLPEVIEPGG KTLARLRL*

20

Computer analysis of this amino acid sequence gave the following results:

25 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF119 shows 93.7% identity over a 175aa overlap with an ORF (ORF119a) from strain A of *N. meningitidis*:

			10	20	30	40	50	60
30	orf119.pep	MIYIVLFLLAVVAVVANMYQENQYRKVKRVDQFGHSKDKALLNSXTSHVRDGKPSSGGSVM						
	orf119a	MIYIVLFLLAVVAVVANMYQENQYRKVKRVDQFGHSKDKALLNSXTSHVRDGKPSSGGSVM	10	20	30	40	50	60
35	orf119.pep	MPKPQPAVKKTAKPQDPXMRNLQEQQDAVYIAKQKQAKASPFTKTEIETALEESGIIGNSAH	70	80	90	100	110	120
	orf119a	MPKPQPAVKKTAKSQQPDMRNQLQEQQDAVYIAKQKQAKASPFTKTEIETALEESGIIGNSAH	70	80	90	100	110	120
40	orf119.pep	TVPSEPQTGHSSATKPADASAKPAPVPPQTPAKPLITLKLKSLKVELSWFDWFVRIDFISY	130	140	150	160	170	
	orf119a	TVPSEPQTGHSSAPKPADAPAKPVPVPPQTPAKPLITLKLKSLKVELSWFDWFVRFDFISYIALTE	130	140	150	160	170	180
45	orf119a	AKELHALPRLSNRCRQYQIVGCTMDDHQPIAEPIPGIRYQAFIGVIGIQAWSRRNGLASQELS	190	200	210	220	230	240

The complete length ORF119a nucleotide sequence <SEO ID 527> is:

50	1	ATGATTCTTACA	TGCTGACTTGT	CCTCCGGCGC	FTCTCTGGCG	TTGTGCGCTA
	51	CAATATGTTAT	CAGSSAACAC	AATAACGSSAA	AAAATGCGGC	GACAGTGTGG
	101	GGCAGCTTCGA	CAAAGTAGGCC	CTGCTGCACAA	GCAAAACCG	CCAGTGCTGCC
	151	GAGGCCAAAC	CCTGCGCCGG	GCCAGTGTAC	ATGCGGAAAC	CCCAACCGG
	201	GTTGCAAAAG	ACGGCAAAAT	AGGAGGACCG	CCCCATGGAT	AACCTTGAGAG
	251	AGGAGGATG	CTCTCTGATC	GCCAAGCAGA	AAACAGCGAA	ACCTGCCCAGG
	301	TTCTAAACCC	AAATCGAAC	GGCCCTTTGAA	AAAGGGCGA	TTATCGCAAA
	351	CTTCGGCCAC	ACCCCTTCG	AAACCCCAAC	CGCACATTC	GGCAACAAAC
	401	CTTCCGGCGC	GGCGGCAA	CCCTTGTCG	TTCCGAAAC	GGCGCCAAAG
	451	CCCGCTTAAT	GGCTGAGAAC	GTCGTCGAAG	GTGCGATGTC	CTGTGGATGAA
	501	CGTGGCTTC	GACTCTCAT	CTTATATGCT	GGCTGACGAA	GCGAAAGAACAA
	551	TGGCAACCTG	GGCGGCCCT	TCCAAGCGGT	CGCGCTTACCA	GATGTGCGGAA
	601	TGGCACCTAC	ACGGCATTT	CGCAAGATTC	GAACCCATTC	CGGGATCGCC

651 CTATCAGGCA TTTATCGTCC GTATTCAAGG AGTCAGGCCG AACCGACTTG
 701 CCTCCGAGGA AGAACTCTCC GCAATTCAACC GCCAGGTGGA TGCATTGCGA
 751 CACAGCATGG CGCGCTCAGAC GCTTCACACC GACCTTGCCTG CCTTTATCGA
 801 ACTGGCTTCC GCACITGAGCC CATTTCTGGC GCGGGCTGAC CAGACTATCG
 851 CCATCCATTG GGTTCGGCC ACCAGATCA GGGCCGTAGA ACTCGCTTCC
 901 CGCGTAACGG CGCTCGTTT CGTTTGGAA GACGACGGG CGTCCACTA
 951 TACCGACACC TGCGGCTCGA CTCATGTC CTCATCTGCG CTCAACACG
 1001 AGCCGTTAAC CAATGCCCCT TTGGACAACC AGTCCTATAA AGCCGTTCACT
 1051 ATCGCTCC ACATCCCGCA TCTTCGGCA GCGGAAAAAA CCTTCGACGA
 1101 TTGTTATG GATTGGCCG TACCCCTCTC CGGCCAGTG AACCTGAATC
 1151 TGCTCAACGG CAAATGGAA GAACITGCG CCCAATGGCT CAAAGACCTG
 1201 CGCACTTATG TATTGGCTCC TCACTCCGAG ATGTCCTAAAG TCCGGTATCGA
 1251 ACCGGCGGC AAAACCCATC TGCGCTGTG CTCCCTAA

This encodes a protein having amino acid sequence <SEQ ID 528>:

15 1 MIYIVLFLAA VLA VVAYNNY QENQYRKVKR DQFGHSDKDA LLSNKTSHVR
 51 51 DCKPSGCFMV MFKPQFQAVKK TAKSQQPAMR NLQEQQDAVII AKQKQAKASP
 101 101 FKTETIETAL ESGIIGNSAH TVEEPQTGHS APFKFADAPAK PVEVFQPTFKA
 151 151 PLTTLKELSE VELPFWVFRF DFISYIALTE AKEHLHPLR SNRCRYQTVG
 201 201 CTMDHFQIA EPIFGIRYQA FIVGQIAVSR NGLASQELS AFNRQVDAFA
 251 251 HSMGCGTLHT DLAIFEVAS ALDAFCARVD QTIAIHVLSP TSISGVELRS
 301 301 AVTGVGFVLE DDGAFHYTT SGSTMFSCS LNNEPFTNAL LDNQSYYKDFS
 351 351 MLIDPHSPA GEKTFDDELM DLA VRLSQLQ NLNLVNDKME EVSTQWLKD
 401 401 RTYVILARQSE MLKVKGIEPGG KTALRLFS*

ORF119a and ORF119-1 show 98.6% identity in 428 aa overlap:

	10	20	30	40	50	60
25	orf119a.pep	MIYIVLFLAA VLA VVAYNNY QENQYRKVKR DQFGHSDKDA LLSNKTSHVR	: : : : : : : : : : :			
	orf119-1	MIYIVLFLAA VLA VVAYNNY QENQYRKVKR DQFGHSDKDA LLSNKTSHVR DQGPSSGGPV	: : : : : : : : : : :			
30		10 20 30 40 50 60				
	orf119a.pep	MPKPQPAVKKTAKS QDPAMRN LQE QDAVYIA KQKQAKA SFKTE IETALEESGII GNSAH	: : : : : : : : : : :			
	orf119-1	MPKPQPAVKKTAKPQDPMRN LQE QDAVYIA KQKQAKA SFKTE IETALEESGII GNSAH	: : : : : : : : : : :			
35		70 80 90 100 110 120				
	orf119a.pep	TVP E PQTGCHS A PKP DA PA K P V F V P Q T P A K P L I T L K E L S K V E L P W F D V R F D F I S Y I A L T E	: : : : : : : : : : :			
	orf119-1	T V S E P Q T G C H S A P K P D A P A K P V P Q T P A K P L I T L K E L S K V E L P W F D V R F D F I S Y I A L T E	: : : : : : : : : : :			
40		130 140 150 160 170 180				
	orf119a.pep	A K E L H A L P R L S N R C R Y Q I V Q C T M D D H F Q I A E P I P G I R Y Q A F I V G I Q A V S R N G L A S Q E E L S	: : : : : : : : : : :			
	orf119-1	A K E L H A L P R L S N R C R Y Q I V Q C T M D D H F Q I A E P I P G I R Y Q A F I V G I Q A V S R N G L A S Q E E L S	: : : : : : : : : : :			
45		190 200 210 220 230 240				
	orf119a.pep	A F N R Q V D A F A H S M G Q G T L H T D L A A F I E V A S A L D A F C A R V E Q T I A I H L V S P T S I S G V E L R S	: : : : : : : : : : :			
	orf119-1	A F N R Q V D A F A Q S M G Q G T L H T D L A A F I E V A S A L D A F C A R V E Q T I A I H L V S P T S I S G V E L R S	: : : : : : : : : : :			
50		250 260 270 280 290 300				
	orf119a.pep	A V T G V G F V L E D D G A F H Y T D T S G S T M F S I C S L N N E P F T N A L L D N Q S Y K G F S M L L D I P H S P A	: : : : : : : : : : :			
	orf119-1	A V T G V G F V L E D D G A F H Y T D T S G S T M F S I C S L N N E P F T N A L L D N Q S Y K G F S M L L D I P H S P A	: : : : : : : : : : :			
55		310 320 330 340 350 360				
	orf119a.pep	G E K T F D D L F M D L V A R L S C Q L N L N L V N D K M E E V S T Q W L K D V R T Y V L A R Q S E M L K V G I E P G G	: : : : : : : : : : :			
	orf119-1	G E K T F D D L F M D L V A R L S C Q L N L N L V N D K M E E V S T Q W L K D V R T Y V L A R Q S E M L K V G I E P G G	: : : : : : : : : : :			
60		370 380 390 400 410 420				
	orf119a.pep					
	orf119-1	370 380 390 400 410 420				
65						

orf119a.pep	KTALRLFSX
orf119-1	KTALRLFSX

5 Homology with a predicted ORF from *N.gonorrhoeae*

ORF119 shows 93.1% identity over a 175aa overlap with a predicted ORF (ORF119ng) from *N.gonorrhoeae*:

The complete length ORF119ng nucleotide sequence <SEO ID 529> is:

20	1	ATGATTATACA	TGCTACTGTG	CCCTCGGCC	GTCCTCGCG	TTGTCGCTT
	51	CAATATGAT	CAGGAAGAAA	AATAACGGAA	AAAATGGCGC	GACATGCTT
	101	GACACTCGG	CAAAGATGGC	CTCTCTAAC	GGAAACACAG	CCATTCGCGG
	151	GACGGGAAAC	CTGGCGGCG	GCCAGTCATG	ATGCCGAAAC	CCCAACGGCG
	201	GTCGTTAAA	CGCCGAAAC	CGGAAAGCTC	CGGCATCGGC	AACCTGCAAA
25	251	AAACAGATGC	CGCTTGATC	GCAAGAGCA	AAACGGCAA	AGCTTCGCGG
	301	TTCTAACAAA	AAATCGAAC	CGCCGTTGAA	GAATTCGCA	TTATTCGCGG
	351	CTTCGGACCG	ACCGTTGCGC	AACCCCAAC	GGACGATTC	GGCGAACGAA
	401	CTTGGCGAACG	GCGCCGAAAA	CCCGTTCGGC	TTCCGAAAGA	GGCGCCGAAAG
	451	CCGGCTGATTA	GGCTGCGGAA	GCTGTCGAG	GTGCGAGTCG	CTCTGTTGCG
30	501	CGTCGCTTC	gACTCTCATC	CCATATTCGC	GTCGACCGCA	GCCAAGAAAGA
	551	TGCAACCCACT	GGCCGGCTTC	tc>AACCGCT	GGCCCTTACCA	GATTTCTCGG
	601	TGCAACCCACT	GGCAAGCATTT	CCAGATTCGC	GAACCCATTC	GGGGCATCCG
	651	CTATCGGCC	TTATTCGTTG	GTATCCAGC	AGTCAGCCCG	AAAGGGACTTC
35	701	CTCTGGACAGG	GAAGACTTCG	GTATCCGAC	GGACGCGGGA	GGGGTTTCCG
	751	CAAAGGTCAG	CGGGCTCAGC	GTCGACACAC	GACGGTTCGG	CTCTTATGCG
	801	ATGTCGCTTC	CGCTACTGGC	CATTCTTCGG	GGGGCTCGAA	GGGACATCCCG
	851	CTACCATTC	GTGTTCTCCG	ACGACATCA	GGGGCTCTAGA	ACTGGGTTCT
	901	GGCTCTAACGG	GGCTGGTTTT	CTTTTGGAA	GACGACGGCG	CTTCCTCAAT
40	951	TACCGCACCC	TGGGGCTCGA	TGGCTTCTTC	ATCTCTGCTC	CTAACACACG
	1001	AGCCGTTTAC	TAATGCTCC	TTGGACATCC	AGCTCTACAG	GGGGCTTACG
	1051	ATGCTGCTCG	ACATCCCGCA	CTCTCGGCCA	GGGGAAAAAA	CTCTCGGAGC
	1101	TTGGTATG	GATTTGGGG	TAGGCTCTGC	CGCTGACTGC	AACTCTGAAT
	1151	TTGGTCAACG	CAAATGGAA	GAAGTTTGC	CCCAATGCGT	CAAAAGCTCT
	1201	GGCTACTTGG	TATTGCGGCC	TCAGCTGGAG	ATGTCGAAAG	TGCGTTATCGG
45	1251	CCCCCTTCCC	AAA>CCCC	CTCCCTCTTC	TTCTCTTA	

This encodes a protein having amino acid sequence <SEQ ID 530>

1 MIYIVLFLAA VLAVVAYNM QENQYRKVKR DQFGHSDKDA LLNSKTSHTVF
 51 DGKPGSGGVF MKEFQPAVKW NALQEQQDAVY AKQGAKQASAK
 101 FETIEITALRE EIGIINNSAH TSVEPQTGH SAKPFDAPAK PVFVEQTPAH
 151 FLITLKELSK WELFWDFRKF DEISYIALTE AKELHALPRN SNRQCYQWV
 201 CTMDHDFQIA EPIFGIRYQA FIVIGQAVSE NLQSACLES AFNRQDADAF
 251 QSMGGQVLT H DLAIFAEVAS ALDAFCARVD CTNIAHEUVP TSISGVELRS
 301 AVTGVGEFLAE DDGAGHYTTD SGSTMFSICL QTNAIEHLPV TL
 351 MLDDGGVSE GEKTFKDTMLD DLVNRSLCSQ NLNLVNDKME EVSTQWLKD
 401 RTBYWLLPROSLL MLLKVCILBGCX KEATLREST

QRE119ng and QRE119-1 show 98.4% identity over 428 aa overlap.

orf119ng	10 20 30 40 50 60
	: : : : : : :
orf119-1	10 20 30 40 50 60
	: : : : : : :

		70	80	90	100	110	120
5	orf119ng	MPKPQPAVKKPAKQFQDSAMRNLQEQA 70					
	orf119-1	MPKPQPAVKKTAKPQDPAMRNLD 70					
10	orf119ng	130 TVSEPQTGHSA 70	140 PKPADAPAKPV 80	150 PQTPQF 90	160 PKPLITLKEL 100	170 SKVELPWFD 110	180 FDFISYIALTE 120
	orf119-1	TVSEPQTGHSA 130	PKPADAPAKPV 140	PQTPQF 150	PKPLITLKEL 160	SKVELPWFD 170	FDFISYIALTE 180
15	orf119ng	190 AKELHALPRLSN 70	200 RCRYQIVGCT 200	210 MDDHQFIA 210	220 EPIP 220	230 GIRYQA 230	240 FIVGQAVSRNG 240
	orf119-1	AKELHALPRLSN 190	RCRYQIVGCT 200	MDDHQFIA 210	EPIP 220	GIRYQA 230	FIVGQAVSRNG 240
20	orf119ng	250 AFNRQRDA 70	260 FAQSNGQ 260	270 LHTD 270	280 LAFF 280	290 EVASAL 290	300 DAPCARV 300
	orf119-1	AFNRQRDA 250	FAQSNGQ 260	LHTD 270	LAFF 280	EVASAL 290	DAPCARV 300
25	orf119ng	310 AVTGVGFV 70	320 LEDDGA 320	330 FHYDTSG 330	340 STMFSCS 340	350 LNNEPFTN 350	360 ALLDNQSY 360
	orf119-1	AVTGVGFV 310	LEDDGA 320	FHYDTSG 330	STMFSCS 340	LNNEPFTN 350	ALLDNQSY 360
30	orf119ng	370 GEKTFDDLF 70	380 MLDLA 380	390 VRLSQLN 390	400 LNLN 400	410 LNVNDM 410	420 EVSTQW 420
	orf119-1	GEKTFDDLF 370	MLDLA 380	VRLSQLN 390	LNLN 400	LNVNDM 410	EVSTQW 420
35	orf119ng	429 KTALARLF 70					
	orf119-1	KTALARLF 429					

Based on this analysis, including the presence of a putative leader sequence in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

45 Example 64

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 531>

1	..GC CGGGCAGC	GCAC GGAAGA	TTC TCTCATG	AACA ACAGCG	ACAC ATCRG
51	GCAGATGTC	GAAR CACCA	CGCG TACGAT	GAAG CTGCTG	ATTTCCTCCA
101	TCCG CCTGTAT	TTCA TTGGTA	GT CGCG GCGA	T CGCG GTGAT	GAAC ATCATG
151	CTCG CTGCC	TTAC CGACCC	CAC CRACCA	AT CGCC ATAC	CG ATGCCAAT
201	CG CGCG CGCG	CG CGCAATA	TT Ty GCAGCA	GT TT TGATT	GAG CGCG GT
251	TAAT CTGCGT	CAT CGCG GT	TT GG TGCGCG	TGG GTT GTGC	CG CGCG CGTC
301	AG CTGCGTGT	TCA ATCATT	TG TAAC CGAC	TT CGCG ATGG	AC ATTC CGCG
351	CAT GT CGCGT	AT CGCG CGCG	T CGCG GTTGTC	GAC CGGAAATC	GCG CAT CGCGT
401	TCGG CTTTAT	GC CTCC AAT	AA RG CA CGCA	AA CTCA ATCC	GAT AGAC GCA
451	TT GG CA CGAG	ATT G			

This corresponds to the amino acid sequence <SEQ ID 532; ORF134>:

1	..ARHG TEE	MNN SNSX IRQIV	ESTT GTM KLL	I SSI ALIS LV	VGG VGM NIM
51	LV SVT ERKE	I G I R M AIG AR	G N IX Q X F L I	E A V L I C V I G G	L V G V G L S A A V
101	S L V I N H F V T D	F P M D I S A M V	I G A V A C S T G I	G I A F C F M P A N	K A A K I N P I D A
151	L A Q D *				

Further work revealed the complete nucleotide sequence <SEQ ID 533>;

This corresponds to the amino acid sequence <SEO ID 534; ORF134-1>:

```

1   MSVQALIHK RNSLILMII LIGTASVVSY VAIQNSGKK LIEPDISSTG
51  NTISI FSPRG FGDGRSNGKII TLITDADAKII AQQSYASAT WMTSGGTLT
101 YRNTDLTASL VGEVGEEQF RGLKLLETGR FDENDWKEDA QVVIQDNQVW
151 DKLFDASDPL GKTILFRKRP LTIVIGVKMND ENAAGNSVDSL MLNSPTTYVM
201 HQTGEGSTHN SITVKWIKND NTQVAEKGLT DLKLRKAGTE FDFNMSNDL
251 RQTVESTGTG MKLIISSAL ISLVVGGIVG NMILMVSLTE RTKELIRGMA
301 IJGARRSLQF QFLIEVTTT VIGGLVGVGL SAAVSLSVNE FVTDFPMDIS
351 AMSVIGAVAC STGIGTAFGF MPANKANAK PIDLACDAD

```

35 Computer analysis of this amino acid sequence gave the following results:

Homology with the hypothetical protein o648 of *E. coli* (accession number AE000189)

ORF134 and o648 protein show 45% aa identity in 153aa overlap:

40	Orf134:	2	RUGTEDFPMNNNSDXIRQIESTGTGCMXXXXXXXXXXXXVVGIGVNIMLWSVTERTKEI	61
			RHG +DFF N D + + VE TT T++	VVGIGVNIMLWSVTERT+EI
	o648:	496	RHGGKPFETWNNDGVLKTVTTRTLQLFLTLVAVISLVNMGIGVNIMLWSVTERTIEI	555
	Orf134:	62	GIRMAIGAARGNNIXQQFLIEAXXXXXXXXXXXXXXXXXXXXXXFNHFVTDFFMDISAMSVI	121
			GIRMA+GAR ++ QQLIEA	F+ + + S +***
45	o648:	556	GIRMAVGARASLDVLQQLFEEAVILCLVNGGALGITLSLLIAFTLQLFLPGWIGFSPLAL	615
	Orf134:	122	GAVACSTGIGIAFGFMPANKAAKLNPIDALAQD	154
			A CST GI FG++PA AA+L+P+DALA++	
	o648:	616	LALFCSTVGILFGLWLPARNAARLDPDALARE	648

50 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF134 shows 98.7% identity over a 154aa overlap with an ORF (ORF134a) from strain A of *N. meningitidis*.

		10	20	30
55	orf134.pep	ARHGTEDFFMNSDXIRQIVESTTGTMKLL		
	orf134a	GESHTNSITVKIDNANTQVAEKGLTDLLKARHGTEDFFMNSDXIRQIVESTTGTMKLL	210 220 230 240 250 260	

-313-

	orf134.pep	ISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMAIGARRGNIXQQFLIEAVLICWIGG
	orf134a	ISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMAIGARRGNILQQFLIEAVLICWIGG
5		270 280 290 300 310 320
	orf134.pep	100 110 120 130 140 150
	orf134a	LGVGLSAAVSLVFNVHFVTDFFPMDISAMSIVGAVACSTGIGIAFGFMPANKAAKLNPIDA
10		330 340 350 360 370 380
	orf134.pep	LAQDX
15	orf134a	LAQDX

The complete length ORF134a nucleotide sequence <SEQ ID 535> is:

1	ATGCGGTGC	AAGCAGTATT	GGCGCACAAA	ATGCGTTCGC	TTCTGACGAT
5	GCTCGCATC	ATCATCGGT	TGCGCTTCGGT	TGCTCTCCGT	GTGCGCATTTG
10	101	GCAACGGTGC	GCAGAARAAA	CTTCTTGAAAC	ACATCAGTTG
15	151	ACACCATCA	GCATCTTC	AGGGCCCGGG	TTCGGGCACA
20	201	CAGGTAATAA	ACCCCTTCCC	TAGAGCAGCAG	AAAAAAATCATC
25	251	GTCAGCTTGC	TTCCGCCAACG	CCCATGACTT	CGAGCGCCGG
30	301	TACCGAATA	CCGACCTGAC	CGCTCTTTC	TACGGTGTGG
35	351	TTTCGACCTG	CGCCGGCTGA	AGCTGAAAC	GGGGGGCCTG
40	401	ACGATGTAA	AGAAGACGGC	ATGGTGTGTCG	TCTACGACCA
45	451	GAAACAATCT	TTGGCGACTC	GGATCCGTTG	GGTAAACCCA
50	501	TTGACCTCA	TCGGCGTGA	GGAAAGAACG	GGAAACGGCTT
55	551	TGCGCAATTG	CGACGCTCTG	ATGCTTTGGG	CCCCCTATAC
60	601	CACCAATCA	CGAGGCGAGG	CCACACCAAAC	TCACATCACCG
65	651	AGAACATGCC	AAATACCCAGG	TTGCGCARAA	TCAAAATCTAA
70	701	AAGCGCGGCA	CGCGACCGGA	GATTTCTICA	TGAACAAACAG
75	751	AGGCAGATAG	TCAAGAACCA	CACGGCTAACG	CGACAGCATC
80	801	CATCGCTCTG	ATTCATTTGG	TAGTCGGGG	CATCGCGGTG
85	851	TGTCGGTGC	CTTATCGAG	CGCACCAAAAG	AAATCGGCAT
90	901	ATCGCGCGG	GGCGCGGCAA	TATTTTCGAG	ACGGATGGCA
95	951	GTTATCTGC	GTCACTCGGG	GTTTGTGTCG	CTGGGGTTTG
100	1001	TCACGGCTGC	GTTCATCAT	TTGTCACCG	TCCGGCCGG
105	1051	GCCATCTCC	TCATCGCGGC	GGTCGCTGTG	GGACATTCTC
110	1101	GTTGGCTCTT	ATGCCCTCA	TCGACCGGAA	TGCGCATCTGC
40	1151	CATTGGCCCA	GGATTTGA	CTAACCTCA	CCGATAGATG

This encodes a protein having amino acid sequence <SEQ ID 536>:

1	MSVQAVLAHK	MRSLLTMLGI	IIGIASVVS	VALGNGSQKK	ILEDISIGTNTISIFPGRG
5	51	NTISIIFPGRG	FDERSRSGRK	TLDIADAKII	AKQSYVASAT
10	101	YRNIDLTASL	YGVGEGQYFDV	RGLKLETGRL	FENDVKEDA
15	151	DKLFDPSDPL	GRTILKFRKR	LTVIGVMRKD	ENAFGNSDVL
20	201	HQITGEHTH	SITVKIRDNA	NTQVAEKGLT	MLSPYTTV
25	251	RQIESTTGTG	MRLLSSIAL	ISLVVGIGIV	NMIMLVSVTE
30	301	IGARGRNLIQ	OPLIBAVLIC	VIGGLVGVGL	RTKEIGIRMA
35	351	AMSVYIAGAC	STGIGIAFGF	SAAVSLVFNH	EVDTPFMDIS

50 ORF134a and ORF134-1 show 100.0% identity in 388 aa over overlap:

	orf134a.pep	MSVQAVLAHKMRSLLTMLGI	IIGIASVVS	VALGNGSQKK	ILEDISIGTNTISIFPGRG
	orf134-1	MSVQAVLAHKMRSLLTMLGI	IIGIASVVS	VALGNGSQKK	ILEDISIGTNTISIFPGRG
55	orf134a.pep	FGDRRSRGIKTLT	IIDAKIIAK	QSYVASAT	PMTSGGTLYRNTDLT
	orf134-1	FGDRRSRGIKTLT	IIDAKIIAK	QSYVASAT	PMTSGGTLYRNTDLT
60	orf134a.pep	RGLKLETGRL	FDENDVKEDA	QVYVIDQN	KDRLFADFSDPLGKTI
	orf134-1	RGLKLETGRL	FDENDVKEDA	QVYVIDQN	KDRLFADFSDPLGKTI
65	orf134a.pep	ENAFGNSDVL	MLSPYTTVM	HQITGEHTNS	ITVVKIDNANTQVAEKG
	orf134-1	ENAFGNSDVL	MLSPYTTVM	HQITGEHTNS	ITVVKIDNANTQVAEKG

	orf134a.pep	DFFMNNNSDSIROIESTTGTMKLLISSIALISLVLVGGIGVMNIMLVSVTERTKEIGIRMA : : : : : : : : : : : : : 5 orf134-1 DFFMNNNSDSIRQIVESITGTMKLLISSIALISLVLVGGIGVMNIMLVSVTERTKEIGIRMA orf134a.pep I GARRGNILQQFLIEAVLICVIGGLVGVLSSA9SLVFNFVTDFFPMDISAMS9IGAVAC orf134-1 I GARRGNILQQFLIEAVLICVIGGLVGVLSSA9SLVFNFVTDFFPMDISAMS9IGAVAC 10 orf134a.pep STG9IAAFGMPANKAAKLN9PIDALA9QDX orf134-1 STG9IAAFGMPANKAAKLN9PIDALA9QDX
		<u>Homology with a predicted ORF from <i>N.gonorrhoeae</i></u>
15	orf134 shows 96.8% identity over a 154aa overlap with a predicted ORF (ORF134.ng) from <i>N. gonorrhoeae</i> :	
20	orf134.pep GESHTNSITVKIKDNANTRVAEKGELAKR9HTEDFFMNNNSD9RQMVESTTGTMKLL	30 : : : : orf134ng
25	orf134.pep ISSIALISLVLVGGIGVMNIMLVSVTERTKEIGIRMAI GARRGNILQQFLIEAVLICVIGG : : : : orf134ng ISSIALISLVLVGGIGVMNIMLVSVTERTKEIGIRMAI GARRGNILQQFLIEAVLICIIIGG	90 : : 324
30	orf134.pep LVGVGLSSA9SLVFNFVTDFFPMDISAMS9IGAVAC9STG9IAAFGMPANKAAKLN9PIDA : orf134ng LVGVGLSSA9SLVFNFVTDFFPMDISA9IGAVAC9STG9IAAFGMPANKAAKLN9PIDA	150 : 384
	orf134.pep LAQD 154	
	orf134ng LAQD 388	

The complete length ORF134ng nucleotide sequence <SEQ ID 537> is:

1	ATGTCGGTGC AAGCA9TATT GCGC9ACAAA ATGCGTTGC TTCTGACCAT	
51	GCTCGGCATC ATCATCGTA TC9GTGCTG TGTCTCCGTC GTGCCGGCTGG	
101	GCACAGGTC GCAGAAAAAA ATCTCGAACG ACATCGACTG GATGGGGACG	
151	AAACACATCA GCATCTTCCC CGGCGGCGC TTCGCGGACAA GGCGCAGCGG	
201	CAAAATCAAA ACCCTGACCA TAGACGACG CAA9AATCATC GCCAACAAA	
251	GCTAGTTGC CTCCGCACAG CCCATGACTG CGACCGCGCG CACGCTGACC	
301	TACCCCAATA CGGACGCTGC CGCTCTCTTG TACCGTTG TG GCGAACATA	
351	TTTCGACGTC CGCGCGCTGA AGCTGGAAAC GGGCGCGCTG TTGATGAGA	
401	ACGAGTGTAA AGAAAGCGG CAAGTCGTC TCATCGACCA AAATGTCAAA	
451	GACA9ACTCT TTGCGGACTC GGATCGGTG G9TAA9ACCA TTGTTGTCAG	
501	GGAA9GCCCC TTAGCGCTCA TC9CGCTGAT GAAAAGACG GAAAAGCCTT	
551	TC9CGCAATTG CGACGTTGCTG ATGCTTTGAT CGCCCTATAC GACGGTGTG	
601	CACCAAAATCA CGACGGAGGA CCAAC9AAC CTCATCACCG TCA9A9TC9A	
651	AGAC9ATGCG AATA9CCC9G TTGCGGAAAAG AGG9CTGCC GAGCTGCTCA	
701	AAGCAGCGCA CGGCAGGCAA GACTCTTTA TGAACACAGC CGACAGCATC	
751	AGGCGAGATG CGAACAGAC CACCGCTACG ATGAGCTGC TGATTTCTC	
801	CATCGGCGCTG ATTTCATG TGACTGGGGG CATCGGTGTC ATGAA9ATTA	
851	TGCTGGTGTG CGTACGCGG CGACCA9AAG AAATGCGCAT AC9GATGCGA	
901	ATCGGC9CGC CGGC9CGGCAA TATTTGCA9 CAGTTTTGTA TTGAGGCGST	
951	GTTAATCTGC ATCATCGAG GCTTGGTGG CGTAGGTTG TCCGCGCCCG	
1001	TCAGCGCTGT GTCA9ATCAT TTGTAACCG ATTCCCGAT G9ACATTTG	
1051	GGCGCATCGS TTATCGGCGC GGTCGCGCTGT TGACCGGAA TC9CGCATCGC	
1101	GTTCGCGCTT ATGCGCTGCCA ATAAGCGACG CAA9ACTCAAT CGCATAGATG	
1151	CATTGGCGCA GGATGTA	

This encodes a protein having amino acid sequence <SEQ ID 538>:

1	M9V9AVLAHK M9S9L9L9L9I L9I9AS9VSV VAL9NG9Q9K I LED9I9SM9T	
51	NT9I9FPGRG FGDR9SGK9K TL9T9D9K9I AKQ9SYVAS9T PMT9SS9GLT	
101	YRN9T9DSL Y9G9E9Y9FDV R9L9K9L9T9G9L9R9L9D9A99 K9W9ID9Q9V9K	
151	DKLFAD9SDPL GKT9L9FR9K9L9R9L9T9V9M9R9D9 ENA9FG9S9D9L9 ML9W9P9T9V9M	
201	H9Q9T9GE9H9T S9T9V9K9D9A9 N9T9V9A9E9G9L9 ELL9K9R9H9T9E D9F9M9N9S9D9I	
251	R9Q9M9V9E9T9T G9T9G9C9T9G9T9G9A9 T9G9C9G9T9G9A9 T9G9C9G9T9G9A9	
301	I GARRGNILQ Q9F9L9E9A9L9C9 I IGG9L9V9G9L9 S9A9V9S9V9F9H F9V9D9P9M9D9S	

351 AASVIGAVAC STGGIAFGF MPANKAAKLN PIDALAQD'

ORF134ng and ORF134-1 show 97.9% identity in 388 aa overlap:

	orf134ng	MSVQAVLAHKMRSLLTMLCIIIGIASVSVSVALGNGSQKKILEDISSMGNTNTISIIFPGRG : : : : : : : : : : : : : 5 orf134-1 MSVQAVLAHKMRSLLTMLCIIIGIASVSVSVALGNGSQKKILEDISSIGNTNTISIIFPGRG : : : : : : : : : : : : orf134ng FGDRRSGKIKTLTIDDAKIIAKQSYYVASATPMITSSGGTLTYRNNDLTAASLYGVGEQYFVD : : : : : : : : : : : : 10 orf134-1 FGDRRSGRKIKTLTIDDAKIIAKQSYYVASATPMITSSGGTLTYRNNDLTAASLYGVGEQYFVD : : : : : : : : : orf134ng RGLKLETGRLFDENDVKEDAQVVVIDQNVKDKLFADSDPLGKTLTFLFRKRPLTVIGVNMKD : : : : : : : : orf134-1 RGLKLETGRLFDENDVKEDAQVVVIDQNVKDKLFADSDPLGKTLTFLFRKRPLTVIGVNMKD : : 15 orf134ng ENAFGNSDVLMLWSPYTVMHQITGEHSNTSITVVKIKDNANTRVAEKGLDELLKARHGT : : orf134-1 ENAFGNSDVLMLWSPYTVMHQITGEHSNTSITVVKIKDNANTVQAEGKLDTLLKARHGT : : orf134ng DFFMNNNSDSIRQMVESTGTMKLLISSIALISLVVGIGVMNIMLVSVERTERTKEIGIRMA : : orf134-1 DFFMNNNSDSIRQIVESTGTMKLLISSIALISLVVGIGVMNIMLVSVERTERTKEIGIRMA : : orf134ng IGARRGNILQQFLIEAVLCIIGLGVGVGLSAAAVSLVFNHFVTDPFMDIASVIGAVAC : : 25 orf134-1 IGARRGNILQQFLIEAVLCIIGLGVGVGLSAAAVSLVFNHFVTDPFMDIASVIGAVAC : : orf134ng STGIGIAFGMPANKAAKLNPIDALAQDX : : orf134-1 STGIGIAFGMPANKAAKLNPIDALAQDX : :
--	----------	---

30 ORF134ng also shows homology to an *E.coli* ABC transporter:

```

sp|P75831|YBJ2_ECOLI HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN YBJ2 >gi_5
(AE000189) 0648; similar to YBAH_HAEIN SW: P45247 [Escherichia coli] Length = 648
Score = 297 bits (753), Expect = 6e-80
Identities = 162/389 (41%), Positives = 230/389 (58%), Gaps = 1/389 (0%)

35 Query: 1 MSVOAVLAHKMRSLTLMLXXXXXXXXXXXXXXLNGSQKKIILEDISSTGNTNTISIIFPRGRG 60
M+ +A- A+A-KMR+LLTM+ +G+ +++ +L DI S+GTNTT ++PG+
Sbjct: 260 MAWRAALANKMRSLTLMLGIIIGIYASIVSVIISVQDAAKQMVLDIARISNTNTIDVYPRGDK 319

40 Query: 61 FGDRRSGKIKTILTDDAKIIAKQSIVVASATPMTSSGGTILTYRNTDLTASLYVGECQQYFDV 120
FGD + L T DD I K Q +VASATP S L Y N D+ AS GV YF+F
Sbjct: 320 FGDDDPQQYQQLAKYKDDLLIAIQKQPWVASATPAVSQNLRLRYNNVDVAASANGVSGDYFNV 379

45 Query: 121 RLGLKLETGRLEPDENDVKEDAQVVVVIDQNVKQFLFAD-SDPLGKTIIFRKRPILTIVGMRK 179
G+ G F++ +VWVBN+D N +LF +D +G+ IL P VIGV ++
Sbjct: 380 YGMFTSEGNTFNQEQLNQRAGQVVLDSNTRPFLPHKRDADVGEVILVGNMPARIVGVAEE 439

50 Query: 180 DENAFGNSDVLMLWSPYTVMVHQTGEHSYNTSITVKIKONANTRVAKGLAEELLKARHG 239
++ FG+S VL -W PY-T++ +G+S NSITV+K++ ++ AE+ L LL RHG
Sbjct: 440 KQSMFGSSKVLRVLPWYSTMSGRVMQGSWLNSITVRVKEGFDSEAEEAQQLTRLLSLRHKG 499

Query: 240 EDFPNNSDLSIRQMVESTWTGMXXXXXXVGGIGVMNIMLVSVERTERTKEIGIRM 299
++ DFF+N D +B + VE TT T++ +VVGIGVMNIMLVSVERTERTKEIGIRM
Sbjct: 500 KDFITWWMDGVLKTEKTTFLQFLTLVAVISLWVGGIGVMNIMLVSVERTERTKEIGIRM 559

55 Query: 300 AIGARGRNLLQFLIEXXXXXXXXXXXXXXXXXXXXXFNHFVTDFFPMISAASVIGAVA 359
+A+GAR ++LQFLIE F+ +S + +A
Sbjct: 560 AVGARASDVLQFLIEAVLVCVLGGALGITLSSLIAFLTQLQFLPWEIGFSPLALLAFL 619

60 Query: 360 CSTGIGIAFPGFMPANKAAKLNPIDALAQD 388
CST GI FG+PA AA+L+P+DAALA+
Sbjct: 620 CSTGIGIAFPGFMPANKAAKLNPIDALAQD 648

```

Based on this analysis, including the presence of the leader peptide and transmembrane regions in the gonococcal protein, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 65

- 5 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 539>:

```

1 ..GGGACCGGG CGATGGCTCT GCTGTGTTTAC CGGGTAACGA T CTGCCCTT
51 GCGCACTGGC GTTACCTCGA GTTACACTCG GTGCATTTTT TTGGCGTAT
101 TLTCCCTTCG GATTGAAAGA GAACCGGATTG CCGTTAACATG CGACGGCGTG
151 CTGCTCTTG GTTTGCGGG CTCGTTATTC TGCTTAACTC CCTCGTCCG
201 CAGGGCTTCAG GAAACCGGG CACTCGCCGG GCTGGCGGG GGCSCGATGT
251 GGCGCTGGGC GTATTGAAAGA GTGGGGCAAC TGCTTTGGC GGGGCAACCC
301 GGCTTGGGGC TGCTGTTTA CCTTCCGCG ACAGGTGTC CGATGTCGTC
351 TTATGCGG ACCTGACCC GCTGGCACAC CCTGTCCTTT CCATCGGGAG
401 TTATCTGTC GTGCATCGCG GTGTCGGCCC TGATGCCCCA ACTGTCGATG
451 ACAGGCTCTC ACAAGTCGG CGACAAMTC ACAGGTTGGCT CGCTTTCTA
501 TATGACCGTC GTTTTCCG CTCTGTCG CGCATTTTTT CTGGGGCAAG
551 ATGGCTTCTG CGACGAAATA CTCGGTATGT GCATCATCAT CTCAGCGGT
601 ATTGTA

```

This corresponds to the amino acid sequence <SEQ ID 540; ORF135>:

```

20 1 ..GTGAMLLFY AVTILPLATG VTLSYTSIF LAVFSFLILK ERISVYTOAV
51 LLGLFAGVUL LLNPFSSRSGQ ETAAALAGLQ GAMSGWYLK VRELSLAGEP
101 GWRUVVFYLSV TGAVAMSSVWA TLTGWHTLSF FSAVILSCIG VSALIAGLSM
151 TRAIKGVDKF TVASLSSYTV FESALSAAFF LGEELFWQEII LGMCIIISAV
201 F*

```

- 25 Further work revealed the complete nucleotide sequence <SEQ ID 541>:

```

30 1 ATGGATACCG CAAAAAAAAGA CATTITAGGA TCGGGCTGGA TGCTGGTGGC
51 GGGGGCCTGC TTTCACATTA TGAACTTATT GATTAAGAG GCAATCGGCAA
101 AATTGCGCT CGGCAGCGGC GAATTGGCTT TTGGGGCCAT GCTGTTTTCA
151 ACCGGTGGGC TCGGGGCTGC CGCCGTAATG CGTCGGACAA CCTTCGGCAC
201 GCGCCATTGG AAAAACCACT TAAACCGGAG TATGGTCGGG ACGGGGGGCGA
251 TGGCTGTGCT GTTTCACCGG GTAAACGATC TGCGCTTGGC CACTGGCGTT
301 ACCCTGAGTT AACACTCGTC GATTTTTTG CGCGTATTTTT CCTTCTCTGAT
351 TTGGAAAGAA CGGATTCCTG TTACACGGA GGCGGTGCTG CCTCTTGGTT
401 TTGCGCGCTG GTGATTCCTG CTTAACTCT CGTTCCGAG CGGTCAAGGA
451 ACGGCGGCAC TCGGGGGCTG GGCAGGGCGC CGGATGTCGG CGCTGGCGTA
501 TTGGAAAGTG CGCGAACCTG CTGGCGGGG CGAACCGGG CGGGCGCTCG
551 TGTTCACCGC TTCCGACAA GTGTGGCGG TGTCTCGGGT TTGGGGCGACG
601 CTGACCGGGC CGCACACCT GTCTTCACCA CGCGCAGTTT ATCTGTCGTC
651 CATGGCGCTG TCGGGCTGTA TTGCGCAACT GTGCGATGAGC CGGCCCTACA
701 AAGTCGGCA CAAAATCAAG GTTGCCTCGC TTTCCTATAT GACCGCTCGT
751 TTTCGCGCTC TGTCGCGGC ATTTTTCTG GCGGAAGAGC TTTCCTGGCA
801 GGAATAACTC GTTATGTGCA TCATCATCTT CAGCGGTATT TGAGCAGCA
851 TCCGCCCCAC TGCCCTTCAA CAGCGGCTGC AATCCCTGTT CGGCCAAAGA
901 TAA

```

- 45 This corresponds to the amino acid sequence <SEQ ID 542; ORF135-1>:

```

40 1 MDTAKKDILG SGWMLVAAC FTIMNVLIKE ASAKFALGSG ELVFWRMLFS
51 TVALGAAAVL RRDXFRTHW KNHLNRSRMVG TGAMLLFYA VHHLPLATGV
101 TLYSTTSIIF AVFSLILKE RISVYTOAVL LLGFAGVULL LNPSFRSGQE
151 TAALAGLAGC AMSGWAYLVK RELSLAGERG WRVVEYLSVT GVAMSSVWAT
201 LTGWHTLSF SAVYSCIGV SALIAQLSMT RAYKGDKFT WASLYSYTVMV
251 FSALSAAFFL GEELFWQEII GNCIILLSGI LSSIRPTAFK QRLQSIFRGR
301 *

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF135 shows 99.0% identity over a 197aa overlap with an ORF (ORF135a) from strain A of *N.meningitidis*:

		10	20	30
5	orf135.pep	GTGAMLLLFYAVTILPLATGVTLSYTSSIF		
	orf135a	STVALGAAAVLRRDTFRTPHWKNHNLNRSMVG		
		50 60 70 80 90 100		
10	orf135.pep	LAVFSFLILKERISVYTQAVVLLGFAGVVL	60 70 80 90 100	80 90 90 90 90
	orf135a	LAVFSFLILKERISVYTQAVVLLGFAGVVL	110 120 130 140 150	120 130 140 150 160
15	orf135.pep	VRELISLAGEPGWRRVVFYLSVTGVAMMSVW	100 110 120 130 140	100 110 120 130 140
	orf135a	ATLGWTLSFPSAVYLSCIGVSALIAQLSM	170 180 190 200 210	180 190 200 210 220
20	orf135.pep	VRELISLAGEPGWRRVVFYLSVTGVAMMSVW	230 240 250 260 270	230 240 250 260 270
	orf135a	ATLGWTLSFPSAVYLSCIGVSALIAQLSM	290 300	290 300

30 The complete length ORF135a nucleotide sequence <SEQ ID 543> is:

1	ATGGATACCG	CAAAAAAAGA	CATTTAGGA	TGGGGCTGGA	TGCTGGTGGC
51	GCGGCCGC	CTTACGCTT	TGAACGTATT	GATTAAGAGG	GCATCGCCGA
101	AATTGGCCCT	CGGCACGGGG	GTATTGGCT	TGGGGCAT	GCTGTTTCA
151	ACCGTTCGGC	TGGCGGCTGC	CCGGCTATTG	CGTCGGGACA	CTTCCCGCAC
201	GCCCCATGG	AAAACCCAT	TAACCCGCA	TATGGCTGG	ACGGGGCGGA
251	TGCTGCTGT	GTTTAACCGG	GTAACGCAT	TGGCTTITGGC	CACCGGGCTT
301	ACCTCTGAGT	ACACCTCTC	GATTTTTTG	GCGGTATTTT	CTTCTCTGTAT
351	TTGAAAGAA	GGGATTTCCC	TTTACACGGA	GGCGGTGCTG	CTCCCTGGTT
401	TTGGCGGCGT	GGTATTGCTG	CTTAATCCTC	CGTTCGGCAG	CGGTCAAGAA
451	ACGGGGCAGC	TGGCGGCGGC	GGCGGGGGGC	GGCATGTTCCC	GCTGGGGCTA
501	TTTAAAGATG	CGGCACACTG	CTTTCGGGG	CGAACCCGGC	TGGGCCCTCG
551	TGTTTACCT	TTCCGTGACA	GCTGTGGCGA	TGTCATCGGT	TGGGGCGACG
601	CTGACCGGGT	GGCACACCCCT	GTCCCTTCCA	TGGCAAGTTT	ATCTGCTGT
651	CATCGCGTGT	TCCGGCGCTGA	TTGCCGCAACT	GTGGATGAGC	CGGGCGTACA
701	AACTCGCGA	CAAATTCAGC	GTTCGCTCCG	TTTCCTATAT	GACCTCGTT
751	TTTCGCGTC	TGTCGCGCG	ATTTTCCTG	GGCGAAGAGC	TTTTCTGGCA
801	GGAAATACTC	GGTATGCGA	TCATCATCCT	CAGCGGTATT	TTGAGCAGCA
851	TCCGCCAAC	TGCCCTCAA	CAGCGGCTGC	AATCCCTGTT	CCGCCAAAGA
901	*				

50 This encodes a protein having amino acid sequence <SEQ ID 544>:

1	MDTAKKDILG	SGWMVLAAC	FTIMNVLIKE	ASAKFALGSSG	ELFWRMLFS
51	TVALGAAAVL	RDRDFTRPHI	KNHLNRSMVG	TGAMLLFYA	VTHPLATGV
101	TLSYTSSI	AVFSFLILKE	RISVYTOAVL	LLGEFAGVULL	LNPSEFRSGQE
151	TAALAGLAG	AMSGKAYLKV	RELSLAGEPF	WRVVFYLSVT	GVAMMSVWAT
201	LITGWHTLSF	SAVYLSCIGV	SALIAQLSMT	RAYKVGDKFT	VASLSSYMTVV
251	PSALSAAFFL	AEELFWQEIL	GNCIIILSGI	LSSIRPTAFK	QRLQSLFRQR
301	*				

ORF135a and ORF135-1 show 99.3% identity in 300 aa overlap:

60	orf135a.pep	MDTAKKDILGSGWMLVAAACFTIMNVLIKEASAKFALGSGELVFWRMLFSTVALGAAVL
	orf135-1	MDTAKKDILGSGWMLVAAACFTIMNVLIKEASAKFALGSGELVFWRMLFSTVALGAAVL

	orf135a.pep	RRTDTFRTPHWNHLNRSNVTGAMLLLFFYAVTHLPLATGVTLSYTSSIFLAVFSFLILKE : : : : : : : : : : : : 5 orf135-1 RRDXFRTPHWNHLNRSNVTGAMLLLFFYAVTHLPLATGVTLSYTSSIFLAVFSFLILKE
	orf135a.pep	RISVYTQAVVLLLGFGAGVVLLLNPSFRSQCETAALAGLAGGAMSGWAYLKVRRELSLAGEPG : : : : : : : : : : : : 10 orf135-1 RISVYTQAVVLLLGFGAGVVLLLNPSFRSQCETAALAGLAGGAMSGWAYLKVRRELSLAGEPG
	orf135a.pep	WRVVFYLSVTGVAMSSVVWATLTGWHTLSPFSAVYLSCLIGVSALIAQLSMTRAYVKGDKF : : : : : : : : : : : 15 orf135-1 WRVVFYLSVTGVAMSSVVWATLTGWHTLSPFSAVYLSCLIGVSALIAQLSMTRAYVKGDKF
	orf135a.pep	VASLSYMTVVFSALSAAFFLAEELFWQEILGMCIILSGLILSSIRPTAFKQRQLQSLFRQR : : : : : : : : : : orf135-1 VASLSYMTVVFSALSAAFFLGEELFWQEILGMCIILSGLSSIRPTAFKQRQLQSLFRQR

Homology with a predicted ORF from *N.gonorrhoeae*

ORF135 shows 97% identity over a 201aa overlap with a predicted ORF (ORF135ng) from

20 *N.gonorrhoeae*:

	orf135.pep	GTTGAMLLLFFYAVTXLPLATGVTLSYTSSIF	30
	orf135ng	STVTLGAAAVLRRDTFRTPHWNHLNRSNVTGAMLLLFFYAVTHLPLATGVTLSYTSSIF	335
25	orf135.pep	LAVFSFLILKERISVYTCAVLLLGFGAGVVLLLNPSFRSQCETAALAGLAGGAMSGWAYLK : : : : : : : : orf135ng LAVFSFLILKERISVYTCAVLLLGFGAGVVLLLNPSFRSQCETAALAGLAGGAMSGWAYLK	90
30	orf135.pep	VRELSSLAEGPGWVVFYLSVTGVAMSSVVWATLTGWHTLSPFSAVYLSCLIGVSALIAQLSM : : : : orf135ng VRELSSLAEGPGWVVFYLSVTGVAMSSVVWATLTGWHTLSPFSAVYLSCLIGVSALIAQLSM	150
35	orf135.pep	TRAYVKGDKFVVASLSYMTVVFSALSAAFFLGEELFWQEILGMCIILSAAV : : orf135ng TRAYVKGDKFVVASLSYMTVVFSALSAAFFLGEELFWQEILGMCIILSAAV	506

An ORF135ng nucleotide sequence <SEQ ID 545> was predicted to encode a protein having amino acid sequence <SEQ ID 546>:

1	MPSKEFKRHH	LRTASFQFGHH	LHHFHHQVKGG	CQIIGFGIHI	FPTLHPLPAQG
51	IILDIQQLGLFR	IDFAAALAVYR	RTQVDFIHTV	IDGIASDQAF	SEVVCLILRL
101	NLGHFTDTHL	IAQAKRRIAD	FGNIRPMRRN	EAKTFCRFCR	FDGIDGIGHD
151	FRQQGHHINRL	APGKDCRNKG	RDKVFFEHTRN	JNQVCLEKTN	CSARKIKFRR
201	QKQAKTHTS	LAARFTIRPS	LSQKPFMMDTA	KKDILGSGWM	LVAACACTVM
251	NNLVIKEASAK	FALGSCELV	WRMLFSTVIL	AAAVALRDRD	FRTHPHWNHL
301	NRSNVTGAM	LLLFYAVTHL	PLTTGVTLS	TTTIIPLAVFS	FLILKERISV
351	YTQAVVLLGF	AGVAVLNLPS	FRSCQEPAL	AGLAGGAMSG	WASYLKVRELS
401	LAGEPGWRV	FYLSATGVAM	SSVWATLTGW	HTLSPFSAVY	LSGIGVSALEI
451	AQCLSMTRAYK	VGDKFTVASL	SYMTVVFSAL	SAAFFLGEEL	FWQEILGMCI
501	TIISAA*				

Further work revealed the following gonococcal sequence <SEQ ID 547>:

50	1	ATGGATACCG	CAAAAAAGA	CATTTTAGGA	TCCGGCTCTGA	TCTGGTGCGC
	51	GGCGGCCTGC	TTCACCGGTTA	TGAACGTTATT	GATTAAGAG	GCATCGGCAA
	101	AATTTCGCCCT	CGGCAAGCGC	GAATTGGCTT	TTTGGCGCAT	CGTGTITTC
	151	ACCGCTTACCG	TCGGTGTGCG	GGCCGTATTG	GGCCGCGACA	CTTCCCGCAC
	201	GGCCCCATTGG	AAAACACACT	TAACACCGCG	TATGGTCGGG	ACGGGGCGCA
	251	TGCTGCTGCT	GTTTTACCGCG	GTAAACCGCAT	TGGCCTTTGAC	AACCGGGCGTT
	301	ACCTCTGAGTT	ACACCTCGTC	GATTTTtttg	GGCGTATTTT	CTTCCCTGT
	351	TTGGAAAGAA	CGGATTCCTG	TTTACACGGA	GGCGGTGCTG	CTCTTTGGTT
	401	TTGGCCGGCT	GCTATTGCTG	CTTATACCTT	CGTTCGGCAG	CGTCAGGAA
	451	CGGGCGGCAC	TGGCGGCGCT	CGGGCGGCAC	GGCATGTCGCG	CTGGGGCGTA
	501	TTGGAAAGTG	CGGGCAACTG	TTTGGCGGG	CGAACCCGGC	TGGCGCGTCG
	551	TGTTTACCT	TTCCCGAACCC	GGCTGTGGCGA	TGTGCTCgt	ttggcgacg
	601	Ctgacccggct	ggCACccct	GTCCCTTccca	tccggcagttt	ATCtgtCGGG

651 CATCGGGGTG tccgcgtgA TTGCCCAaCT GtcgtatGcg cCGCgttaca
 701 aaGTCGGCGA CAAATTACAGC GTTGCCTCGC ttttccataAT gaccgttGTC
 751 TTTTCGGCC TGTCTGGCG ATTTTTCTG ggccggaaagc tttTctggCA
 801 GGAACTACTC GGTATGTCGA TCATTAatcc CAGCGGCAATT TTGAGCAGCA
 851 TCCGGCCCAT TGCCCTCAA CAGCGGCTGC AAGCCCTT CCGCCAAAGA
 901 TAA

This corresponds to the amino acid sequence <SEQ ID 548; ORF135ng-1>:

1 METAKKDILG SGWMLVAAC FTVMVNVLIKE ASAKFALGSGELVFWRMLFSTVTLGAAAVL
 51 TVTLGLAAVL RRDTRFTFHW KHNHLNRSVMVG TGAMLLFYA VTHPLTTGV
 101 TLSYTTSIIF AVFSFLLEK RISVYQAVL LIGFAGVVLL LNPSFRSGE
 151 PAAALAGLACG AMGSWAVLKV RELSLAGEB RG NWVVFYLALSAT GVAMSSWAT
 201 LTGWHTLSFP SAVYLSGIVG SALIQLSMT RAYKVGDKFT VASLSIMTVV
 251 PSALSAAFFL GEELFWQEIL GMCIIILSGI LSSIRPIAFK QRLQALFRQR
 301 *

15 ORF135ng-1 and ORF135-1 show 97.0% identity in 300 aa overlap:

orf135ng-1.pep	MDTAKKDILGSGWMLVAACFTVMVNVLIKEASAKFALGSGELVFWRMLFSTVTLGAAAVL orf135-1 MDTAKKDILGSGWMLVAACFTIMVNVLIKEASAKFALGSGELVFWRMLFSTVALGAAAVL
orf135ng-1.pep	RRTDFRTPHWNHNLNRSVMVG TGAMLLFYA VATHPLTTGVLSYTTSIFLAFLVSFLILKE : orf135-1 RRDTRFTPHWNHNLNRSVMVG TGAMLLFYA VATHPLATGVTLSYTTSIFLAFLVSFLILKE
orf135ng-1.pep	RISVYQAVVLLIGFAGUVLLLNPSPRSQCPEPAALAGLAGGAGMSGWAYLKVRSLSLAGEPG orf135-1 RISVYQAVVLLIGFAGUVLLLNPSPRSQCPETAALAGLAGGAGMSGWAYLKVRSLSLAGEPG
orf135ng-1.pep	WRVVFYLSATGVAMSSVWATLTGHWTLFSPSAYVLSGIVGSALIAQLSMTRAYKVGDKFT : orf135-1 WRVVFYLSVTGVAMSSVWATLTGHWTLFSPSAYVLSICIGVSALIAQLSMTRAYKVGDKFT
orf135ng-1.pep	VASLSYMTVFSALSAAFFLGEELFWQEIILGMCIIILSGLLSSIRPIAFKQRLQALFRQR orf135-1 VASLSYMTVFSALSAAFFLGEELFWQEIILGMCIIILSGLLSSIRPTAFKQRLQSLFRQR

35 Based on this analysis, including the presence of several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 66

The following DNA sequence was identified in *N.meningitidis* <SEQ ID 549>:

40 1 ATGAGCGGG GTATAGCCGT CCTGGCTCG TTCCCGAGA TAATCCGAGT
 51 TTGGGGACAA CTGTTGGCGA AAATCGTCAA TACAGTTCCG GCACATCGGA
 101 TGCTCTTCCA GATTTGGGG ATGTTCTTTC TCTTCATACAA CCAGCAATAT
 151 CTGGCGGGGA TCGCCGAAT CGATTCCTCA TGCGGCATCG TTGTCGGGTG
 201 GCTCTCTTC CGTCATCTCG CGCGGCGATTC CCTGTATGGT AAAGCCGGCG
 251 TAGGGGATTC CGTTCACAC GACATCCAG TGCGCTGATGT CGTCAACCG
 301 AACGCAACAT CTTTGCGCTT GTTCGACATG GGTCACTTGC CCGGGTTCAT
 351 TGTTCAGCA ACCGTAATAA TAAAGACCGT CAAAATAAT ATCGTCGATC
 401 CACATATGTT CGCAAAATTTC CGCGCTCTTC CGCTCTTGGG AAAAAGGGAC
 451 TTGACCATG CAAARAAATCA AGGGGGAAAT AATGGGGCGG CGTCTCCAAA
 501 AAAGCTGGC CAAAAATATAT TTGAATGTTT TACGGGGCGG TTGTCGGCA
 551 CGGTTCAGC GTTGCTCTGC CTGTTCTACAA TAAATAATGA CGGAAATGCC
 601 CATATCATTC GCTCTCTAAC CGTTCAGGGTA TCTGTGTCGA CCTTACTGCG
 651 GCTTCTGcC kTCGGCATCC GATTCGGGAT TGAAAAGTTC mrrwyATTCG
 701 GAATAG

55 This corresponds to the amino acid sequence <SEQ ID 550; ORF136>:

1 MKRRIAVFVLPQIIRVLQG LLPKIVNTVP AHRMLFQIFG MFFFFIHHQY
 51 LPGLAEIDSP CGIVPGALPF RHLFAHCLYK KAAVGDVAH EHFVADVVNR

-320-

101 NANAFALFDI GQFAKXIVQH TVNIKTVKIN IVDPHMFANF AVFAVLEKRD
151 FDHGKIQGGN NAAAFPKKLA PKIFECFTGA FVGTVTREVFC LFYIINDGIA
201 HHSAPQRVRY LEFAFGCGFLP SADSDSLKS XXSE*

Further work revealed the complete nucleotide sequence <SEQ ID 551>.

5	1	ATGATGAAAGC	GCGCTTACGG	CCTGTTCTGC	CTGTGTCCTCCG	AGATAAACCGG
	51	AGTTTTTGGA	CAACTGTTGC	CGAAATAGCT	CAATACAGTC	TGAACTTCTG
	101	GGATGCTCTT	CGAACATTGTT	GGGATGTTCT	TTTCTTCTCAT	ACCAACGCAA
	151	TATCTGCCGG	GGATCGCGGA	AATCGATTC	CCATCGGGCA	TCTGTTCTGC
	201	TGGCGCTCTC	TTCCGTCATC	TGGCCGGGCA	TTGGCTGTAT	GTAAAGACCGC
10	251	CGCTGGGGGA	TGGCGTGTCA	CAGCACATC	CAGTCGCTGT	TGGCTGCTAT
	301	CGGAAAGCGA	ACGCTTCTTC	CTTGTGTCAC	ATTGGTCAGT	TGGCGGGTTG
	351	CATTGTTGAC	CAACACGGCA	ATATAAAGACG	CGTCGAAATAA	ATATACTGGCT
	401	ATCACATACAT	GTGGCAAAAT	TTGGCGCTCT	TCGGCCTGCT	GGAAAAAGGG
	451	GACTTGTGACC	ATGGCAAAAC	CAAAGGGCGGA	AAATAAGCGG	GGCCGGCTTC
	501	AAAAGAGCTC	GCGCCAAAAA	TATTGGATG	TTTACCGGG	GCGTCCTTC
	551	GCACGGTTTA	CGGGTTCTGC	TGGCTTCTTCT	ACATATAAAG	TGGCAAGGAT
	601	GCCCCATCAT	CTGTCCTCTCA	ACGTGTCAGG	TATGGTTGTT	CACCTTACTG
	651	CGCTGTTCTT	CGCTTCGGCT	CGGATTGGGA	TTTGAAAAGT	TCCAAATATT
	701	CGGAATAG				

20 This corresponds to the amino acid sequence <SEQ ID 552; ORF136-1>:

1 MMKRR1AVFV LFPQIIRVLG QLLPK1VNTV PAHRLMFOIF GMFFFFIHQQ
5 YLPGIAEIDS PCGIVFGALL PHRLPACRHL GKAAGVDAVA HEHPVADVVN
10 RNAAANALPFD IKGQAFGVFVQ HTVN1VTKVI NIVDPHMFAFN FAVFAVELEKK
15 DFDHGKIQGG NNAAAFFPKKL APKIFK1VCFQ AFVGTVYRFV CLEYIINDGIV
20 AHHSAPQRVY LFYAPFYCGFL PSASDSDLKS SKYE*

Computer analysis of this amino acid sequence gave the following results

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF136 shows 71.7% identity over a 237aa overlap with an ORF (ORF136a) from strain A of *N. meningitidis*:

The complete length ORF136a nucleotide sequence <SEO ID 553> is:

1 ATGATGAGACG GCGCTTACGG CGTCTTCGTC TCTGCAATGC AAUAAAATC
 51 GATTGGGGCA CAACTGTGTC CGAAATAGTG CAATACAGTC CGGCAGCATC
 101 GGATGATGTTG CCAAGATNTG GGGATGTTCTT TTCTTCATC AACAGCACAG
 151 TACCTGCCCTT GGATCAGCGGA AATCCATTTC AGATCGGCCA TGCTTGTCG
 201 TAGCGTCATC TTCCGCTATC NGTCACCCCA TGCTGCTGTA GTGAAACGCC
 251 CGCTTGGGA TGGCTTGGCA CAGCGAACATC CAGTCCTGTA GTGCTAACAG

-321-

301 CGGAACGCAA ACGGCTTCGC CTTGTTGCGC ATTGGTCAGT TCGCCGGGTT
 351 CATTGTTTCAG CAGCCAAATAA ATGTAAAGAC CGTCAAAATAA AAATACGTCG
 401 ATCCACCATAT GTTCGCAAAAT TTCCGCNTCT TCAGCGTCTT GAAAAAAAGG
 451 GCTTGTGACCA TGCGCAAAATC TAAGGNGNNNA NNAGATGCGGC GGCGTTCCCA
 501 AAAAACGCTG CGGCCAAAAAT ATTGGTAAAGT TTTCGCGGGC CTTCGCGCG
 551 CACGGTTTAC CGGTTTGTCTC CGCTGTTCTCA CATAATAAAT GACGGAAATCG
 601 CCCATCATAT CTGCTCTCTCA ACGGTGTCAGG TATCTGTTTG CACCTTACTG
 651 CGGGTTTCTG CCTTCGGCAT CGGATTGCGA TTTGAAAAGT TCCAAATATT
 701 CGGAAATAG

10 This encodes a protein having amino acid sequence <SEQ ID 554>:

1 MMKRRIAVNFV LLMQKIRILG QLLPKIVNTV PAHRLMLFQXF GMFFFFIHQY
 51 YLPGIAEIDS FCGIVFGPLL FRHXRTHCLV GKAAGVNAVA HEHPVADVVN
 101 RNNAAFALED IIGCFAGFIVQ HAINVKTVKI NIVDPHMFPAN FAXFAVLEKR
 151 ALTMAKSXXX XMRRRSQKSS RQKYLNVLRA RSPARFTGLS ACST**MTE
 201 PIISAPQRVRV YLFAPYFCGL PSASDSDLKS SKYSE*

ORF136a and ORF136-1 show 73.1% identity in 238 aa overlap:

	10	20	30	40	50	60
orf136a.pep	MMKRRIAVFLVLMQKIRILGQLLPKIVNTVPAHRLMLFQXFMFFFFFIHQYQLPGIAEIDS					
orf136-1	MMKRRIAVFLFPQIIRVLGQLLPKIVNTVPAHRLMLFQIFGMFFFFFIHQYQLPGIAEIDS	20	30	40	50	60
	70	80	90	100	110	120
orf136a.pep	PCGIVFGTLLFRHXRTHCLYGKAAGVNAVAHEHPVADVVNVRNNANAFALFDIGQFAGFIVQ					
orf136-1	PCGIVFGALLFRHLPAHCLYGKAAGVCAVHEHPVADVVNVRNNANAFALFDIGQFAGFIVQ	70	80	90	100	120
	130	140	150	160	170	180
orf136a.pep	HAINVKTVKINIVDPHMFPANFAXFAVLEKRALTMASKXXMRRRSQKSRSQKYLNVLRA					
orf136-1	HTVNVIKTVKINIVDPHMFPANFAVFAVLEKRDFFDHGKIQGGNNAAAFPKKLAPKIFECFTG	130	140	150	160	180
	190	200	210	220	230	
orf136a.pep	R---SPARFTGLSACSTXXMTEESIISAPQRVRYLFAPYFCGFPLPSASSDLKSSKYSEX					
orf136-1	AFVGTVYRFVCLFYIINDGIAH---SAPQRVRYLFAPYFCGFPLPSASSDLKSSKYSEX	190	200	210	220	230

40 Homology with a predicted ORF from *N.gonorrhoeae*

ORF136 shows 92.3% identity over a 234aa overlap with a predicted ORF (ORF136ng) from *N.gonorrhoeae*:

45	orf136.pep	MKRRIAVFLFPQIIRVLGQLLPKIVNTVPAHRLMLFQIFGMFFFFFIHQYQLPGIAEIDS	59
	orf136ng	MMKRRIAVFLLMQKIRILGQLLPKIVNTVPAHRLMLFQIFGMFFFFFIHQYQLPGIAEIDS	60
	orf136.pep	PCGIVFGALLFRHLPAHCLYGKAAGDVAHEHPVADVVNVRNNANAFALFDIGQFAGFIVQ	119
50	orf136ng	PGGIVFGTLLFRHLPAHCLYGKAAGDVAHEHPVADVVNVRNNANAFALFDIGOSAGFIVQ	120
	orf136.pep	HTVNVIKTVKINIVDPHMFPANFAVFAVLEKRDFFDHGKIQGGNNAAAFPKKLAPKIFECFTG	179
	orf136ng	HTVNVIKTVKINIVDPHMFPANFAVFAVLEKRDFFDHGKIQGGNNAAAFPKKLAPKIFECFTG	180
55	orf136.pep	AFVGTVYRFVCLFYIINDGIAH---SAPQRVRYLFAPYFCGFPLPSASSDLKSSXXSE	234
	orf136ng	AFAGTVYRFVCLFYIINDGIAH---SAPQRVRYLFAPYFCGFPLPSASSDLKSSKYSE	235

The complete length ORF136ng nucleotide sequence <SEQ ID 555> is:

60 1 ATGATGAAAGC GGCCTATAGC CGTCTTCGTC CTGCTCATGC AGAAAATCCG
 51 GATTTGGGA CAACCTTGGC CGAAAATCGT CAATACAGTT CGGGCACATC

101 GGATGCTCTT CCAAATTTTC GGGATGTCTT TTTCTCAT ACACCGGCAA
 151 TACCTGCCG GGGATGCCGA AATCGATTCC CCAAGGGGTA TCGTGTTCGG
 201 TAGCTCTCTC TTCCGCTATC AATGGGGCA TTGCGCTGAC GTAAAGCCG
 251 CGGTAGGGGA TGCCGTTGCA CAGAACACATC CAGTGGCTGA TGTGCGAAC
 5 301 CGGAAGCCAA ACGGCTTGCG CTGGTGCAC ATGGTCACT CGCGCGGGT
 351 CATTGTCATC CACACCTTAA ATATAAAGAC CGTCAAAATAA ATATTCGTG
 401 ATCCACATATG TTTCGCAAT TTCCGCGTGT TCGCGCTTT GGAAAAAAGG
 451 GACTTGGACG ATGGCAAACTT CAAAGGGGAA AATAATGCGG CGGGGTTC
 501 AAAAAGCTC GGGCCAAAGG TATTGTAATG TTTTACGGGC CGGTGCGCG
 551 GCAGGGTTTA CGCGTTCGTC TGGCTGTTCT ACATAAATAAA TGACGGAAATC
 601 GGCATCATATA CTGCTCTCTCA ACGTGTACGG TATCTGTTG GACCTTACCG
 651 CGGTGTTCTA CCTCCGGCAT CGGATTGCGGA TTTGAAAAAGT TCCAATATT
 701 CGGAATAG

This encodes a protein having amino acid sequence <SEQ ID 556>:

15 1 MMKRRIAVFV LIMQKIRILG QLLPKIVNNTV PAHRLMLFQIF GMFFFFIHRQ
 51 51 YLPGIAEIDS PGGIVFGFTLL FRHLSAHCLY GKAAVGDAVA HEEHPVADVAN
 101 101 RNNAAFAFLD ICQSGAFIVQ HTVNIRKTVKI NIVDPHMIFAN FAVFVAFLEKR
 151 151 DFDHKGQDG NNAAAFPKQL APKVFEFCFG AFAGTIVRFV CLFYIINDGI
 201 201 AHHTPAQRVR YLFAPYRQFL PPASDSDLKS SKYSE*

20 ORF136ng and ORF136-1 show 93.6% identity in 235 aa overlap:

orfl36ng	MMKRRIAVFV LIMQKIRILG QLLPKIVNNTV PAHRLMLFQIF GMFFFFIHRQ YLPGIAEIDS
orfl36-1	MMKRRIAVFV LIMQKIRILG QLLPKIVNNTV PAHRLMLFQIF GMFFFFIHRQ YLPGIAEIDS
25 orfl36ng	PGGIVFGTLL FRHLSAHCLY GKAAVGDAVA HEEHPVADVAN RNANAAFLD DIGQSAGFIVQ
orfl36-1	PGGIVFGTLL FRHLSAHCLY GKAAVGDAVA HEEHPVADVAN RNANAAFLD DIGQSAGFIVQ
30 orfl36ng	HTVNIRKTVKINIVDPHMIFAN FAVFLEKRDHFHDGKIQGGNNAAFPKPLAKPKFECFTG
orfl36-1	HTVNIRKTVKINIVDPHMIFAN FAVFLEKRDHFHDGKIQGGNNAAFPKPLAKPKFECFTG
35 orfl36ng	AFAGTIVRFVCLFYIINDGI AHHSAPQRVRYLFAPYRGFLPPASDSLKSSKYSEX
orfl36-1	AFAGTIVRFVCLFYIINDGI AHHSAPQRVRYLFAPYRGFLPPASDSLKSSKYSEX

Based on the presence of the putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 67

40 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 557>:

1 ATGGAAAATA TGGTAACGTT TCCAAAATC AGACCGCTTT TGGCAATCGC
 51 CGCCGGCGCG TTGCTGCCG CC.TGCGGAC GGGGGGAATAT AATGCTGTCC
 101 GCAAGCGGT GAAACCGCC AAACCCGCC CAGTGGTGGG TTTCGCACTC
 151 GTGGGGGGGT CATTCAAGG ATTGGCCCAT GTAGGTATTA TTAAGGTTTT
 45 201 GAAAGAAAAGG GTTATTCCTG TGAAAGCTGGT TACCGCACC TCCGCAAGTT
 251 CGATTCGCG CAACCTTTT GCATCGGGTA TGTGCCCCGA CGCGCTCGAA
 301 TTGGAAAGCGG AAATTTTAAAG CAAACCGAT TTGGTGGATT TAACCTGTG
 351 CACCAATGGG TTATCAAG GCGCAAGGCT GCAAAATTAC ATCACACCGAA
 401 AACTCGCGCG CATGCAGATT CAGCAGTTT CCATCAAATT TGCGCC..

50 This corresponds to the amino acid sequence <SEQ ID 558; ORF137>:

1 MENNVTFSKI RPLLAIAAAA LLAXXRTAGN NAVRKPVQTA KPRANVGLAL
 51 GGGASRGFAH VGIIKVLKEN GIPFKVVKVTGT SAGSIVGNLF ASGMPSDFRLE
 101 LEAEILGKTD LVLDLTLSTG FIKGAKLQNY INRKLRGMQI QOFFIKFAA..

Further work revealed the complete nucleotide sequence <SEQ ID 559>:

1 ATGGAAAATA TGGTAACGTT TTCAAAAATC AGACCGCTT TGCAATCGC
 51 CGCCGCGCGC TTGCTTCGCG CCTGGCGCAC GGCGGAAAT AATGCTGCCC
 101 GCAAGCGGTG CAAACCGCCG AAACCGCCG CAGTGGTCGG TTGGCACTC
 151 GGTGGCGCGC CATCTAAAGG ATTGCCCCAT GTAGGTATTAA TTAAAGTTTT
 5 201 GAAAGAAAAC GGTATTCTG TGAAAGTGTTG TACCGGCACA TCGCGAGGTT
 251 CGATTGTCGG CAGCGCTTTT GCATGGGTTA TTGCGCCCGA CGCGCTCGAA
 301 TTGGAGCGC AAATTTTAGG CAAACCGGT TTGGTGGATT TAACCTTGTG
 351 CACCGATGGT TTATCAANAG GCGAAAAGCT GCGAAAATTAC ATCAACCGAA
 401 AAGTCGGGG CAGCGCATTT CAGCAGTTTC CCATCAAAT TGCGCCGTT
 451 GCTACTGATT TTGAAACCGC CAAGGCCGTC GCTTCAATC AGGGGAATG
 501 CGGGGAGGCT GTGGGGCGCT CGCGGCCAT TCCCAATGTG TTGCAACCCG
 551 TTATCATCGG CAGGCATACAA TATGTTGACG GGGTCTGTG CGACGGCGCTG
 601 CCCGTCAGT CGGCCCGCG GCAGGGGGCG AATTTCGTGA TTGCGCTGGA
 651 TATTCGGCGC CGTCCGGCGA AAAACATCGC CCAAGGTTTC TTCTCTTATC
 701 TCGATCACAGC GTGCAACGTA ATGAGCGTTT CTGGTGTGCA AATGAGTGT
 751 GGGCAGGGG ATGTTGAGT CAAACCGCAG GTTTTGGATT TGGGTGCA
 801 CGGGGATTC GATCAGAAAA AACCGCCCAT CGGGTTGGGT GAGGGGGCAG
 851 CACGTCGGCG ATTGCGTGA ATCAAACGCA AACTGGCGC ATACCGTTAT
 901 TGA

20 This corresponds to the amino acid sequence <SEQ ID 560; ORF137-1>:

1 MENMVTFSKIRPLLAIAAAA LLAACGTAGN NAVRKPVQTA KPAAVVGLAL
 51 GGASKGFAH VGIILVKLEN GIPVKVWTGT SAGSIVGSLF ASGMPSPDRLE
 101 LEAEILGKTD LVLDLTLSTSG FIKGEKLQNY INRKVGGRQI QQPPIKFAAV
 151 ATDFETGKAV AFNGNQAGA VRASAIINV PQVPIIGRHT YVDGLLSPV
 201 PVSARRQGA NEVIAVDISA RPKNISQGF FSYLDQTNLV MSVSAALQNEI
 251 QGADWVVKPQ VLIDLGVVGF DQKKRAIRLG EEAARAALPE IKRKLAAHY
 301 *

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

30 ORF137 shows 93.3% identity over a 149aa overlap with an ORF (ORF137a) from strain A of *N. meningitidis*:

	10	20	30	40	50	60
orf137.pep	MENMVTFSKIRPLLAIAAAA LLAACGTAGN NAVRKPVQTA KPAAVVGLAL	GGGASKGFAH VGIILVKLEN GIPVKVWTGT SAGSIVGSLF ASGMPSPDRLE	LEAEILGKTD LVLDLTLSTSG FIKGEKLQNY INRKVGGRQI QQPPIKFAAV	ATDFETGKAV AFNGNQAGA VRASAIINV PQVPIIGRHT YVDGLLSPV	PVSARRQGA NEVIAVDISA RPKNISQGF FSYLDQTNLV MSVSAALQNEI	QGADWVVKPQ VLIDLGVVGF DQKKRAIRLG EEAARAALPE IKRKLAAHY
35 orf137a	MENMVTFSKIRPLLAIAAAA LLAACGTAGN NAVRKPVQTA KPAAVVGLAL	GGGASKGFAH VGIILVKLEN GIPVKVWTGT SAGSIVGSLF ASGMPSPDRLE	LEAEILGKTD LVLDLTLSTSG FIKGEKLQNY INRKVGGRQI QQPPIKFAAV	ATDFETGKAV AFNGNQAGA VRASAIINV PQVPIIGRHT YVDGLLSPV	PVSARRQGA NEVIAVDISA RPKNISQGF FSYLDQTNLV MSVSAALQNEI	QGADWVVKPQ VLIDLGVVGF DQKKRAIRLG EEAARAALPE IKRKLAAHY
	10	20	30	40	50	60
orf137.pep						
40 45 orf137a	70 VGIILVKLEN GIPVKVWTGT SAGSIVGSLF ASGMPSPDRLE	80 EAEILGKTD LVLDLTLSTNG	90 100 110 120			
orf137a	VGIILVKLEN GIPVKVWTGT SAGSIVGSLF ASGMPSPDRLE	EAEILGKTD LVLDLTLSTNG	70 80 90 100 110 120			
	130	140	149			
45 orf137.pep	FIKGAKLQNYYINRKLRGMQIQQPPIKFAA					
orf137a	FIKGAKLQNYYINRKLRGMQIQQPPIKFAA	ATDFETGKAV AFNGNQAGA VRASAIINV PQVPIIGRHT YVDGLLSPV	140 150 160 170 180			

The complete length ORF137a nucleotide sequence <SEQ ID 561> is:

50 1 ATGGAAAATA TGGTAACGTT TTCAAAAATC AGACCGCTT TGCAATCGC
 51 CGCCGCGCGC TTGCTTCGCG CCTGGCGCAC GGCGGAAAT AATGCTGCCC
 101 GCAAGCGGTG CAAACCGCCG AAACCGCCG CAGTGGTCGG TTGGCACTC
 151 GGTGGCGCGC CATCTAAAGG ATTGCCCCAT GTAGGTATTAA TTAAAGTTTT
 55 201 GAAAGAAAAC GATGATTCG TGAAAGTGTTG TACCGGCACA CGCGCAGGTT
 251 CGATAGTCGG CAGCGCTTT CAGATCGGTTA TTGCGCCCGA CGCGCTCGAA
 301 TTGGAAAGCG AAATTTAGG TAAACCGAT TTGGTGGATT TAACCTTGTG
 351 CACCGATGGT TTATCAAG GCGAAAAGCT GCGAAAATTAC ATCAACCGAA
 401 AAGTCGGGG CAGGCAGGTT CAGCAGTTTC CCATCAATT TGCGCCGCTT
 451 GCTACTGATT TTGAAACCG CAGCGCTTC GCTTCAATC AAGGGAAATG
 501 CGGGGAGGCT GTGGGGCGCT CGCGGCCAT TCCCAATGTG TTGCAACCCG
 551 TTATCATCGG CAGGCATACAA TATGTTGACG CGGGTCTGTC CGACGGCGCTG

601 CCCGTCAGTG CGCCCCGGCG GCANGNNNN NATNTCGTGA TTGCGCGTGA
 651 TATTTCCGCC CGTCGCAGCA AAAACATCAG CCAAAGGCTTC TTTCCTTATC
 701 TGATGACAG GCTGAACTA ATGAGCGTTT CCGCGTGC AAAATGAGTTG
 751 GGGCAGCGCG AGTGTGTTAT CAAACCGCG GTTTTGATT TTGGTGCAGT
 801 CGCGGATTTC GATCAGAAA AACCCGCAT CC CGGGTGGGT GGAGGAGCG
 851 CACGTGCCGC ATTGCCTGAA ATCAAACGCA AACTGGCGGC ATACCGTTAT
 901 TGA

This encodes a protein having amino acid sequence <SEQ ID 562>:

1 MENNVTFSKIRPLLAIAAAA LLAACGTAGN NAARKPVOTA KPAAVVGLAL
 51 GCGAGSKGFGH VGIIKVLKEN GIPVVKVUTGT SAGSIVGSLF ASGMSPDRLE
 101 LEAELIGKTD IVDLTLSTSG FKGEKLQNY INRKVUGBRI QOFPIKFRAV
 151 ATDFETGKAV AFNQGNAAQCA VRASAIIPNV FQPVIIGRHT YVDGGLSQPV
 201 PVSAAARRXXX XXVIADISA RPSRNISQGF FSYLDQTILNV MSVSALQNEL
 251 QOADVVVIKPQ VLDDLGAVGGF DQKKRAIRLG EEAARAALPE IKRKLAAYRY
 301 *

ORF137a and ORF137-1 show 97.3% identity in 300 aa overlap:

10	orf137a.pep	MENNVTFSKIRPLLAIAAAA LLAACGTAGNNAARKPVOTA KPAAVVGLAL LGGGASKGFAH
	orf137-1	
20	orf137a.pep	VGIIKVLKENGIPVKVVTGTSAGSIVGSLFASGMSPDRLEAEIYGKTDLVDTLSTSG
	orf137-1	
25	orf137a.pep	FKGEKLQNYINRKVUGRRIQOFPIKFAA VATDFTGKAVAFNQGNAGQAVRASAAIPNV
	orf137-1	
30	orf137a.pep	FQPVIIIGRHTYVDGGLSQPVPSAARRXXXXXIAVDISARPSKNISQGFFSYLDQTLNV
	orf137-1	
35	orf137a.pep	MSVSALQNLQNLQADVVVIKPQVLIDLGVAVGGFDQKKRAIRLGEEAARAALPEIKRKLAAYRY
	orf137-1	

Homology with a predicted ORF from *N.gonorrhoeae*

ORF137 shows 89.9% identity over a 149aa overlap with a predicted ORF (ORF137ng) from

N.gonorrhoeae:

40	orf137.pep	MENNVTFSKIRPLLAIAAAA ALLAAXRTAGNNNAVRKPVOTA KPAAVVGLAL LGGGASKGFAH	60
	orf137ng		60
45	orf137.pep	VGIIKVLKENGIPVKVVTGTSAGSIVGSLFASGMSPDRLEAEIYGKTDLVDTLSTNG	120
	orf137ng		120
50	orf137.pep	FKGEKLQNYINRKLRGMQIQOFPIKFAA	149
	orf137ng		180

The complete length ORF137ng nucleotide sequence <SEQ ID 563> is:

1 ATGGAATAA TGGTAACGTT TTCAAAATC AGATCATTTT TGGCAATCGC
 51 CGCCGCCGG TTGCTTCCGC CCTGCGTAC GCGGGAAAC AATGCCCGCC
 101 GCAAGCGGT AAACCCGCCG CAGTGGTGC TTTGGCACTC
 151 GTGGCGCGG CATCTAAAGC ATTGGCCAT ATAGGAATG TTAAAGTTT
 201 GAAGAAAAAC GTATTCCTG TGAAGGTGGT TACCGGCACA TGGCAAGTT
 251 CGATAGTCGG CAGCCTTTG GCATCGGGTA TTGTCGCCGA CGCGCTCGAA
 301 TTGGAAGCCG AGATTTTAGA TAAACCGAT TTAGTGTGATT TAACCTTGTG
 351 CACCATGTTT TTATCAAAAG CGCGAACAGT GCAAAATTC ATCACCGGA
 401 AAGTGGCGG CAGCGCGATT CAGCAGTTTCCATCAATT TGCCTCCGTT
 451 GCCACTGATT TTGAAACCGG CRAGGGCGTC GCTTTCATC AGGGAAATGC

501 CGGGCAGGGC GTTCGTGCTT CGGCAGGCCAT TCCCAATGTT TTCCAGCCAG
 551 TCATCATCGG CAGGGACAAA TATGTGAGC GCGGTCGTGCG GCAGCCCGTG
 601 CCCGTCAGTG CGCGTCGCGC CGAGGGCGC AATTTCCTGA TTGGCGCTCGA
 651 TATTTCGGCA CCTCCGGAGCA AAAATGTCGG TCAAGGTTTC TTCTCTTATC
 701 TGATGACAG CTCGAACGTC ATGAGCTTTC CGGTGTTGCA AACAGAGTTG
 751 ggycAGGGCGT ATGTTGGATT CAACAGCGag gtTTTGGATT TGGGGTCAGT
 801 CGGGGATTC GATCAGAAAA AGCGGCCCAT CGGGTTGGC GAGGAGGCG
 851 CACGTCCGCG ATTGCGTGA ATCAAACCCA AACTGGGGC ATACCGTTAT
 901 TGA

10 This encodes a protein having amino acid sequence <SEQ ID 564>:

1 MENMVFSKIRPLLAIIAAALLAAGCTAGNNAAARKPVQTA KPAAVVVALAL
 51 GGGASKGPFH IGIIVKVLKSEN GIPVKVVTGTT SAGSIVGSLL ASGMSPDRLE
 101 LEAILELIGRTD LVDLTLSTSG FIRGEKLQNLY INRKKVGRQI QQQPIKFAAV
 151 ATDFETGRAV AFNQNGAQAA VRASAAIPNV FQPVIIGRHK YVDDGLSQPV
 201 PVSAAVRQGQF NFVIAVDISIA RPSKVNQGQGF FSYLDQTLNV NSVSVLQNEL
 251 GQADVVIKPQ VLDDLGAVGFF DQKRRAIRLG EEAARAALPPE IKRKLLAAYRY
 301 *

ORF137/ng and ORF137-1 show 96.0% identity in 300 aa overlap:

20	orf137ng	MENMVFSKIRPLLAIIAAALLAAGCTAGNNAAARKPVQTA KPAAVVVALALGGGASKGPFH : : : : : : : : : : : orf137-1
		MENMVFSKIRPLLAIIAAALLAAGCTAGNNAAARKPVQTA KPAAVVGLALGGGASKGPFH
25	orf137ng	IGIVKVLKENGIVPKVVTGTSAGSIVGSLLASGMSPPRLEEARIELGKTDLVDTLSTSG : : : : : : : : : orf137-1
		: : : : : : : : : VGIVKVLKENGIVPKVVTGTSAGSIVGSLLASGMSPPRLEEARIELGKTDLVDTLSTSG
30	orf137ng	FIRGEKLQNLYINRKKVGRQIQQFFPKFAAVATDFETGRKAWAFNQGNAGQAVRASAAIPNV : : : : : : : orf137-1
		FIKGEKLQNLYINRKKVGRQIQQFFPKFAAVATDFETGRKAWAFNQGNAGQAVRASAAIPNV
35	orf137ng	FQPVIIGRHKYVDGGLSOPVPSAARRQGANFVIAVDISARPSKVNQGQFSYLDQTLNV : : : orf137-1
		FQPVIIGRHTYVDGGLSOPVPSAARRQGANFVIAVDISARPGKNISQGFFSYLDQTLNV
40	orf137ng	MSVSVLQNELQGADVVIKPQVLDDLGAVGFPDKKKRAIRLGEEAARAALPPEIKRKLLAAYRY : orf137
		MSVSALQNELQGADVVIKPQVLDDLGAVGFPDKKKRAIRLGEEAARAALPPEIKRKLLAAYRY

Based on the presence of a predicted prokaryotic membrane lipoprotein lipid attachment site (underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and

40 *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 68

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 565>:

45 1 ATGTTCTGTT TACAATTCTAG GCTGTTTCCC CCTTTGGGAA CGCCCATGCA
 51 CATCTCTGTTT ACCGGCTCTGC TCAAATGCGT CTCCCTGCTG CGCGCTTTCCT
 101 GTCCTGGACAT GCTGGGAAC CGGGCTGGAC ATCTGGCGGT TTACCTTTTA
 151 AAAGGAAGACG GGGCGGCCAT CGTCGCCCCAT ATGGGGCAGG CGGGTTTGAA
 201 CCCGGACCCC AAAACGGCTCA AAGCGCTTTT TGCGGAAACG GCRAAAAGCG
 251 GTTGGAACT TGGCCCCGGG TTITTCAGAA AACCGGAAGAA CATAGAAACA
 301 ATGTTCTAAG CGGTACACCGG CTGGGAACAT GTGCGAGCAGG CCTTGGACAA
 351 ACAGCAAGGG CTGCTATTC..

This corresponds to the amino acid sequence <SEQ ID 566; ORF138>:

55 1 MFLRLQFRLFP PLRTAMHILL TALLKOLSLL PLSCLHTLGN RLGHLFAYLL
 51 KEDRARIVAX NRQAGLNPPD KTVKA_VFAT AKGGLELAPA FFRKPEDIE
 101 MFKAHVGEWEH VQQALDKHKEG LLF

Further work revealed the complete nucleotide sequence <SEQ ID 567>:

```

      1 ATGTTTGGTT TACAATTCAG GCTGTTTCCC CCTTTGCCAA CGGCCATGCA
      5 CATCTGTG AC CGCCCTCGC TCAAATGCT CTCCCTGCTG CGGCTTTCTT
     10 CTCTGCACAC GCGGCCAAC CGCTGCCAGC ATCTGGCTT TTACCTTTA
     15 AAGGAAGACC CGCGCGCAT CGTCGCCAM ATCGGCCAGG CGGTTTTGA
     20 CCCGGACCC AAAACGCTCA AAGCCGTTTG TGCGGAAACG GCAAAAGGCC
     25 GTTTGGAACT TGCCCCCGG TTTTCAGAA AACCGGAGAA CATAGAACAC
     30 ATGTCAAAAG CGGTACACGG CTGGGAACAT GTGCAAGCAGG TTGGACACA
     35 ACAGCAAGGG CTGCTATTCA TCAGCGCGA CATCGGCCAGC TAGCATTTGG
    10 401 GCGCAGCTGA CATACCGCAAG CAGCTTCCCGT TCCCGCTGAC CGCCATGTAC
     451 AAACCCCGGA AAATCAAAAGC GATAGACAAA ATCATGCGAC CGGGCAGGT
     501 TCGCCGCAAAG GAAAGAACCC CGCTTACACAG CATACAAGGG TCAACACAAA
     551 TCATCAAAGC CCTGGCTGC GCGCAAGCNA CCATCGTCT CGCCGACAC
     601 GTCTCCCTCG CTCAAGAAGG CGGGGAAGGC GTATGGTGG ATTCTCTGG
    15 651 CAAACCTGCG TATCACATGA CGCTGGCGC AAAATGGCA CACCTCAAAG
     701 GCGTGAAGAAC CCTTGTGTTTC ATCTGGCAAC GCGTGCCTGG CGGACAAAGGT
     751 TTGGCAATTGCG ACATCCGCC CGTCCAAAGG GAATTGACCC GGACAAAGC
     801 CCATGATGCC CGCGCAATGCA ACCCGCAATGC CGAATATTTGG ATAGGGCGTT
     851 TTCCGAGCGA GTATCTGTTT ATGTAACACC GCTACAAAT GCGCTAA

```

20 This corresponds to the amino acid sequence <SEQ ID 568; ORF138-1>:

```

      1 MFLRQFLRFP PLRTAMHILL TALLKCLSLIPLSCLHTLGNRLGHAFYLLKEDRARIVAX
      5 KEDRARIVAN MRQAGLNPDP RTVKAVFAEST AKGGLELAPAF RRKPEDIST
     10 MFKAHVIGWEH VQQALDKHEO LLFITPHIGS YDLGGRYIQL QLPFPFLTMAV
     15 KPPKIKAIKD IMQAGRVRGX GKTAPTSIQG VQKIIKALRS GEATIVLPDH
     20 VPSPQEGEG WWVDFGKPA YTMTLAALKL HVKGVKTLFF CCERLPGGK*
     25 FDLIRPVQG ELNGDKHA AVFNRNAEYW IRFPFTQYL MYNRYKMP*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF138 shows 99.2% identity over a 123aa overlap with an ORF (ORF138a) from strain A of *N.*

30 *meningitidis*:

		10	20	30	40	50	60	
	orf138.pep	MFLRQFLRFPPLRTAMHILLTALLKCLSLIPLSCLHTLGNRLGHAFYLLKEDRARIVAX						
	orf138a							
35		MFLRQFLRFPPLRTAMHILLTALLKCLSLIPLSCLHTLGNRLGHAFYLLKEDRARIVAN	10	20	30	40	50	60
	orf138.pep							
	orf138a							
	orf138.pep	MRQAGLNPDPRTVKAVFAETAKGGLELAPAFFRKPEDIEITMFKAHVIGWEHVQQALDKHEG	70	80	90	100	110	120
40	orf138a	MRQAGLNPDPRTVKAVFAETAKGGLELAPAFFRKPEDIEITMFKAHVIGWEHVQQALDKHEG	70	80	90	100	110	120
	orf138.pep	LLF						
45	orf138a	LLFITPHIGSYDLGGRYISQQLPFPLTANYKPPKIKAIKD KIMQAGRVRGKGTAPTSIQG	130	140	150	160	170	180

The complete length ORF138a nucleotide sequence <SEQ ID 569> is:

```

      1 ATGTTTGGTT TACAATTCAG GCTGTTTCCC CCTTTGCCAA CGGCCATGCA
      5 CATGGTGTG AC CGCCCTCGC TCAAATGCT CTCCCTGCTG CGGCTTTCTT
     10 CTCTGCACAC GCGGCCAAC CGCTGCCAGC ATCTGGCTT TTACCTTTA
     15 AAGGAAGACC CGCGCGCAT CGTCGCCAM ATCGGCCAGG CGGTTTTGA
     20 CCCGGACCC AAAACGCTCA AAGCCGTTTG TGCGGAAACG GCAAAAGGCC
     25 GTTTGGAACT TGCCCCCGG TTTTCAGAA AACCGGAGAA CATAGAACAC
     301 ATCTCAAAAG CGGTACACGG GTGGGAACAT GTGCCAGCAGG TTGGACACA
     351 AGACCGAGGG CTGCTATTCA TGAGGGCGG CATCGGCGAC TACGATTG
     401 CGCGACCTCA CATCAGCGAG CAGCTTGGGT TCCCGCTGAC CGCCATCTAC
     451 AAACCGGCCA AAATCAACG GATAGACAAA ATCATGCGAG CGGGCAGGGT
     501 TCGCGCAAAG GCGAAACCG CCCTAGAGG CATACAAAGG GTCAACACAA
     551 TCATCAAAGC CCTCGCTTCG GCGGAAGGAA GCATCGTCT CGCCGACAC

```

-327-

601	GTC COCT CCCC	CTCAAGAAGG	CGGGGAAGGC	GTATGGGTGG	ATTCTTCGG
651	CAACACTGCC	TATACTATGA	CGCTCNGGCC	AAAATTTGCA	CACGTCAAAG
701	GGGTAAACAC	CTCTTITTC	TGCTCGAAC	GCCCTGCCTG	CGGACAAAGT
751	TTCGATTTGC	ACATCGCCC	CTTCCAAGGG	GAATTGAAACG	GGGACAAAGC
801	CCATGATGCC	GGCGTGTCTCA	ACCCGAATCC	CGAATATTGG	ATACCGCGTT
851	TTCCGACGCA	GTATCTGTT	ATGTAACACC	GCTACAAAT	GGCGTAA

This encodes a protein having amino acid sequence <SEQ ID 570>:

1	MFRLOFRLFPPLRTAMHILLTALLKCLSLPLSCLHTLGNRLGHLFYLL
51	KEDRARIVAN
101	MRQAGLNPDPKTVKAVFAETAKGGLELAPAFFRKPKEDIETMFKA
151	VHGWEHVQQALDKHEG
201	LLFTPHIGSYDLGGGRYISQQLPFPFLTAMYKPPKIAIDKIMQAGRVRGKGKTAPTSIQC
251	VKGVKTLFF

ORF138a and ORF138-1 show 99.7% identity over a 298aa overlap:

15	orf138a.pep	MFRLOFRLFPPLRTAMHILLTALLKCLSLPLSCLHTLGNRLGHLFYLLKEDRARIVAN
	orf138-1	MFRLOFRLFPPLRTAMHILLTALLKCLSLPLSCLHTLGNRLGHLFYLLKEDRARIVAN
20	orf138a.pep	MRQAGLNPDPKTVKAVFAETAKGGLELAPAFFRKPKEDIETMFKA
	orf138-1	VHGWEHVQQALDKHEG
25	orf138a.pep	LLFITPHIGSYDLGGGRYISQQLPFPFLTAMYKPPKIAIDKIMQAGRVRGKGKTAPTSIQC
	orf138-1	LLFITPHIGSYDLGGGRYISQQLPFPFLTAMYKPPKIAIDKIMQAGRVRGKGKTAPTSIQC
30	orf138a.pep	VKQIIKKALRSGEATIVLPDHVPSPQEGEGGVWVDFFGKPAYTMTLAAKLAHVKGVKTLFF
	orf138-1	VKQIIKKALRSGEATIVLPDHVPSPQEGEGGVWVDFFGKPAYTMTLAAKLAHVKGVKTLFF
35	orf138a.pep	CCERLPGQQFDLHRRPVQGELNGDKAHDAAVFNRNAEYWIRRFPTQYLEMNYRKKMP
	orf138-1	CCERLPGQQFDLHRRPVQGELNGDKAHDAAVFNRNAEYWIRRFPTQYLEMNYRKKMP

Homology with a predicted ORF from *N.gonorrhoeae*

ORF138 shows 94.3% identity over a 123aa overlap with a predicted ORF (ORF138ng) from *N.gonorrhoeae*:

40	orf138.pep	MFRLOFRLFPPLRTAMHILLTALLKCLSLPLSCLHTLGNRLGHLFYLLKEDRARIVAX	60
	orf138ng	MFRLOFRLFPPLRTAMHILLTALLKCLSLLSLSCHTLGNRLGHLFYLLKEDRARIVAN	60
45	orf138.pep	MRQAGLNPDPKTVKAVFAETAKGGLELAPAFFRKPKEDIETMFKA	120
	orf138ng	VHGWEHVQQALDKHEG	120
50	orf138.pep	LLF	123
	orf138ng	LLFITPHIGSYDLGGGRYISQQLPFPFLTAMYKPPKIAIDKIMQAGRVRGKGKTAPTSIQC	180

The complete length ORF138ng nucleotide sequence <SEQ ID 571> is:

50	1	ATGTTTCGTT	TACAATTCAG	GCTGTTCGCC	CTTTTCGAA	CCGGCATGCA
	51	CAT CCT GTTG	ACCGCCCTGC	TCAAATGCCT	CTCCCTGCTG	TCGCTTCT
101	101	GCTCTCACAC	GCTGGGAAAC	CGCTCGGCAC	ATCTGGCGTT	TTACCTTTTA
151	151	AAGGAGAAC	GGCGCGGCAT	CTFCGGCAAT	ATCGGCAAGG	CGGGTTTGA
201	201	CCCCGACACG	CAGACCGCTCA	AAACCGCTTT	TGGCGAAACG	GCAAAATTCG
251	251	TGTTGAACT	TGGCCCCCGC	TTTTCAAA	AACCGGAAAG	CATCGAAACA
301	301	ATGTTCAAG	CGGTACAGG	CTGGGGACAC	GTGCGACAGG	CTTGGACAA
351	351	GGCGCAAGGG	CTGCTGTTC	TCACGGCGCA	CATCGGCGAC	TACGATTG
401	401	GGCGACGCTA	CATCGCCGCA	CGCTTCCG	TCCACCTGAC	CGGCATGTAC
451	451	AAGCCGCGCA	AAATCAAAACG	GATAGACAA	ATCATCGAGG	CGGGCAGGGT
501	501	GGCGCGCAA	GGCAAAACG	cggccacccgg	catACAAGGG	GTCAAACAAA
	551	tcatcaAGGC	CTGCGCCGG	GGCGAGGCAA	CCAtcATCCT	GGCCGACCCAC

601 GTCCCTTCTC CGCAGGAagg cggCGGCGTG TGGGCGGATT TTTTCGGCAA
 651 ACCTGATAC acATCGACAC TGCGGCCAAA ATTGGCACAC GTCAAAGGGC
 701 TGAACACCT TTGTTTCTGC TGCGAACGCC TGCGGACGG ACAAGGCTTC
 751 GTGTGCACT TCAGGCCCGT CCAAGGGGA TTGACCGGA ACAAGGCCA
 5 801 CGATGCCGC CGATGCAACC GCATACCGA ATATTGGATA CGCCGTTTC
 851 CGACCGAGTA TCTGTTTATG TACAACCGCT ATAAAACGCC GTAA

This encodes a protein having amino acid sequence <SEQ ID 572>:

1 NFRLQFRLFPPLRTAMHILL TALLKCLSL SLSCLHTLGN RLGHHLAFYLL
 10 KEDRARIVAN MRQAGLNPDQT QTVKAVFAET AKCGLELAPLA FPKKPEDIE
 101 MFKAHVGMHEH VQOALDKGEQ LLFITPHIGS YDLGGRYIQL QLPFHILTAMY
 151 KPPKIKAIKD IMQAGRVRSGK GKTA PTFGQG VKQIIKALRA GEATIILPDH
 201 VPSPQGGGV WADFFGKPAY TMTLAKLAH VKGVKTLFFC CERLPDGQGF
 251 VLHIRPVQG LNGNKAHDAA VNRRNTEYWI RRFPTQYL FM YNRYKIP*

ORF138ng and ORF138-1 show 94.3% identity over 299aa overlap:

15	orf138-1.pep	MFRHQFRLFPPLRTAMHILL TALLKCLSL SLSCLHTLGN RLGHHLAFYLL KEDRARIVAN
	orf138ng	MFRHQFRLFPPLRTAMHILL TALLKCLSL SLSCLHTLGN RLGHHLAFYLL KEDRARIVAN
20	orf138-1.pep	MRQAGLNPDQT QTVKAVFAET AKCGLELAPLA FPKKPEDIE TMFKAHVGMHEH VQOALDKHEG
	orf138ng	MRQAGLNPDQT QTVKAVFAET AKCGLELAPLA FPKKPEDIE TMFKAHVGMHEH VQOALDKHEG
25	orf138-1.pep	LLFITPHIGSYDLGGRYISQQLPEFLTAMYKPPKIAKIDKIMQAGRVRVGKGKTA PFTSITQG
	orf138ng	LLFITPHIGSYDLGGRYISQQLPEFLTAMYKPPKIAKIDKIMQAGRVRVGKGKTA PFTSITQG
30	orf138-1.pep	VKQIIKALRSEATIVL PDHVPSPQEGEGVWVDFFGKPAY TMTLAAKLAH VKGVKTLFF
	orf138ng	VKQIIKALRSEATIVL PDHVPSPQEGEGVWVDFFGKPAY TMTLAAKLAH VKGVKTLFF
	orf138-1.pep	CCERLPQGGQFDLHIRPVQGE LNGDKAHDR AFRVN RNAEY WIRRFTQYL FM YNRYKMP
	orf138ng	CCERLPQGGQFDLHIRPVQGE LNGDKAHDR AFRVN RNAEY WIRRFTQYL FM YNRYKTP

In addition, ORF138ng is homologous to htrB protein from *Pseudomonas fluorescens*:

35	gnl PID e334203 (Y14568) htrB [Pseudomonas fluorescens] Length = 253
	Score = 80.8 bits (196), Expect = 9e-15
	Identities = 49/151 (32%), Positives = 79/151 (51%), Gaps = 6/151 (3%)
40	Query: 101 MFKAHVGMHEH VQOALDKGEGLL FITPHIGSYD -LGGRYISQQLPFH LTAMYKPPKIAKID 159
	+ + V G E ++ AL G+G++ IT H+G++ L Y S Q P Y+P+K+KA+D
	Sbjct: 94 LVREVEGLEVLREALASGGKGVGITSHLGNWEVLNHFYCSCQKPI --IFYRPKKLAVD 150
45	Query: 160 IMQAGRVRVGKGKTA PFTGQVQIIKALRSEATIVL PDHVPSPQEGEGVWVDFFGKPA 219
	+++ RV+ K A + +G+ +IK +RG I D P P E G++ FF A
	Sbjct: 151 ELLRKQRVQLGNGKVAASTKEGILSVIKEVRKGQGVIPAD--PEPAESAGIFVPPFFATQA 208
	Query: 220 YAHVTKLAAHVKGVKTLFFCCERLPQGGQF 250
	T + +F RLPDG G+
	Sbjct: 209 LTSKFPVNMLAGGKAVGVFLHALR LPDGSGY 239

50 Based on this analysis, including the presence of a putative transmembrane domain in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF138-1 (57kDa) was cloned in the pGex vectors and expressed in *E.coli*, as described above.

The products of protein expression and purification were analyzed by SDS-PAGE. Figure 14A

55 shows the results of affinity purification of the GST-fusion protein. Purified GST-fusion protein was used to immunise mice, whose sera were used for ELISA (positive result) and FACS analysis

(Figure 14B). These experiments confirm that ORF138-1 is a surface-exposed protein, and that it is a useful immunogen.

Example 69

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 573>:

```

5      1 .. CGGTGGTCGG CCGCGAATC GTGGCGTGTG TTAATGGAAA GTGAAACGTG
51     51 GCATCGCGT TGGATACTT TCAGCGCTCTC GCGCGCGGCC GTGTTATCGCC
101    101 ATGCCCGGCCG TGATGTTTTA GCGCGTTATG GTGCGCCCGG TTGTTGTTTC
151    151 ATGCCCGGCCG TGATGTTTTA GCGCGTTATG GTGCGCCCGG TTGCGGTGTC
201    201 CGCGGGCGTG CTGCTGCTT ATCCGCACTG GACGGCTTCTG TTGCGGTGTC
251    251 TGCCTGGCGAT GTATCGCTG TGCGCGTATG CGTTTGTGCG AAAAGATTT
301    301 TATCATCAGCTT GGATGTCGACT GCGCCCGGGAAT TACCGCAGGG CGCGCGCGG
351    351 TTTCGGGTCGA AACCGCTTTC AGACCGCATG CGCGATCACG TTCCCCCTCT
401    401 TGAACACCGC CGCGCGGCGG GTGCTGACTT TTGCGGGCGGC AACCTGGCTG
451    451 GCGGAAATTG CGCGCGCATT GTTCTCTGCG CGTCCCGGAAT GCGAGACGCT
501    501 GACGACTTTG ATTATAGCCT ATTGGGAGC CGCGGGTAGA GATAATTACG
551    551 CGCGGGCGAT GGTTGCTG..
```

This corresponds to the amino acid sequence <SEQ ID 574; ORF139>:

```

1 .. AWSAGESWRV LMESETWHAV WNTLRFSAAA VYAAAVLVGVV YAAPARRSWA
51     51 MRGLMXPFBM VSPVCVSAGV LLLYFWFTAS LPILLAMVALY LPYPTVAKDV
101    101 LSAWDALPPD YGRAAAGLGA NGFQTACRIT FPLLKPALR GLTLLAATCDV
151    151 GEFAATLFLS RPWQQTLLT IYAYLGRAGE DNYARAMVL..
```

Further work revealed the complete nucleotide sequence <SEQ ID 575>:

```

1 ATGGATGGAC GCGCTTGGGT GGTATGGGGT GCTTTGGCCC TGCTGCCTTC
51 GGCCTTTTTC CGCGTAATGGT TGTGTTGGCG TTGTTGCGGCG TTGGCGGGCGT
101 ATGAGCTTGGT GGCGTGGGGC GGGGTGGTGT CGGATGCGTCA TATGCTCAA
151 CGTTTGGCGT GGACGCGATTC TCAAGGACGG GCACCGTGTG TCTCTGGTGT
201 GCGCTTGGCG TGCTGCTGCTC CTGGTTGGTGT GCGCGCGCGTG CGGTTTCCGG
251 CGCGGGCTTGGT GGTCGCTGGC CTGCTGATTC TGCGCTTGTGT GATGCCAACG
301 TTGGTGGCGGC GCGTGGGGGT GCTGGCCCTG TTGGGGGGGG ACGGGGCTGTT
351 TTGGCGGGGG AGCGAGATA CGCCGATATC GTTGTGTAC GCGAAATGTG
401 TTTCACACTT TCTCTGGTGT GTCAAGGGGG CGTACAGGGG GTTGGTGTCAA
451 TTGGCTGGCG CACGGCTTCGA GACGGCACGG ACCTTGGCG CGGGGGGGTGT
501 CGCGGGCTTGG TGGGACATTC AAATGCGCGT TTGGCGCGCG TTGGCTTGGCG
551 CGCGGGCTGGT CCTGCTGCTC CTGTTATGGT TTCCCGGGT CGGGCTGGCG
601 CTGGCTGCTGG CGCGCAGCGG TTATGCGACCG GTGCGAATGGG AAATTAAACCA
651 TTGGTGTGATG TTGCAACCTCG AATATGGCGT TTGCTTGGTGT CGGGGTGGC
701 TTGTTGTTGGG GTTAAGGGGG CGCGCGAGGT TGCCTGTATGC GTGGTTGGCG
751 AGGGCGCGCG TTTGGAAATAA GCGCGTTTCC CCTGTTGATGC CGTGGCGGCC
801 CGAGCTGGTC GGGGAAATATC TGCTGCTGCC GTTGGGGCGG GGGGGTGTGT
851 CTGGTGTCTG CCTGTTATCG TTGTTGGCAA TTGTTGTGAA AGCTGTGTG
901 CGCGCGCAAT CGTGGCTGTG TTAAATGGAA AGTGAACAGT GCGACGGGGT
951 GTGGAATACT TTGCGCTCTC CGCGCGGGGG GTGTGTTATCG CGGGCGGGTTT
1001 TTGGTGTGGT GTATCGGGCG CGCGCGGGGG GGTGCGCTGTG GATCGCGGGG
1051 CTGGTGTGGT TTGGCGCTTAT GTGTCGCGCG TTGTTGTGTG CGGGCGGGGT
1101 GCTCTGGCTT ATTCGCGCATG CGGACGGCTTC GTTGGCGGTG CTGCTGGCGA
1151 TGTATGGCTT GTGCGCGATG CGGTTTGTGG CAAAAGATGT TTATCAGGCC
1201 TTGGATGACAT CGCGCGGGGG TTACCGCAGG CGCGCGGGGG TTGGGGTGTG
1251 AAACCGCTTT CAGACCGCAT CGCGCATCAC GTTCCCCCTCG TTGAAACCGG
1301 CGTTCGCGGG CGGTCTGACT TTGGCGGGGG CAACCTGGGT CGGGCAATTG
1351 CGCGCGACAT TTGTTCTGTC CGGTCGCGGA TTGGCAGACGG TGACGACTTT
1401 GATTATGCC TATTGGGAC CGCGCGGGTG GAATTAATTAC CGCGCGGGCGA
1451 TTGGTGTGAC ATTCGCTGTG CGCGCGGTGCG CGTGGGTAT TTCTCTGCTG
1501 TTGGACGGGG CGCAAGGGGG AAAACAGACG GAAACGTTTAA
```

This corresponds to the amino acid sequence <SEQ ID 576; ORF139-1>:

```

55 1 MDGRRWWVNG AFALLPSAFL AVMVVAPLNA VAAVDGLAWR AVLSDAYMLK
51 51 RLAWTVEQRA ATCVLVPLG VEVFAWVRLR AIFGRALVLR LLMLPFVMPY
101 101 LVAGVGVLAL FGADGLLWRG RQDTPYLLY GNVFNLFPV LRAAYQQFVQ
151 151 VPAARIQATAR TLGAGAWRFL WDIEMPVLRP WLAGGYCLVF LYCFSGEGGLA
```

201 LLLGGGSRYAT VEVEIYQLVM FELDMAVASV LVWLVLGVT AAGLLYANFG
 251 RRAVSRSKAV PVPMSPPQSV GEYVLLAFAA AVLSVCCCLF LLAIVVKAWS
 301 AGESWRVLM ESETWQAVWNR LRFSAAAVVA AAALGVVYAA AARRSAWMRG
 351 LMFLPPMSP VCVSAGVLLL YPQWTASLFL LLAMYALLAY FFVAKDVLSA
 5 401 WDAIIPPYGR AAAGLGANGF QTACRITFPL LKPALRRGLT LAAATCVGEF
 451 AATLFLSRE P WOTLTTLIYA YLGRAGEDNY ARAMVITLLL AAFAALGIFEF
 501 LDGEGGKQT ETL*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

- 10 ORF139 shows 94.7% identity over a 189aa overlap with an ORF (ORF139a) from strain A of *N. meningitidis*:

	orf139.pep	orf139a	10	20	30
15		OSVGEYVLLAFAAVXSVCLFXLLAIVKWAWSAGESWRVIMESETWQAVWNTXKRFSSAA			
		270 280 290 300 310 320			
20	orf139.pep	YYAAAVLGVVYAAPARISAWMRGLMFXEPFMVSPVVCVSAGVLLLYPQWTASLPLLAMYAL	40	50	60
	orf139a	YYAAAVLGVVYAAPARISAWMRGLMFI_PFMVSPVVCVSAGVLLLYPQWTASLPLLAMYAL	330 340 350 360 370 380	70	80
25	orf139.pep	LAYPFVAKDVLSSADALPDYGRRAAGLGANGFQTA CRITFPLKPAIRRGLTLLAATCV	100	110	120
	orf139a	LAYPFVAKDVLSSADALPDYGRRAAGLGANGFQTA CRITFPLKPAIRRGLTLLAATCV	390 400 410 420 430 440	130	140
30	orf139.pep	GEEFAATLFLSRPEWQTLLTLIYAYLGRAGEDNYARAMVY	160	170	180
	orf139a	GEEFAATLFXSRXEWQTLLTLIYAYXGRAGXNDYARAMVITL LLLAAFA1GXFLLDGGEGG	450 460 470 480 490 500	189	

- 35 The complete length ORF139a nucleotide sequence <SEQ ID 577> is:

```

1 ATGGATGGAC GCGCGGGC GGTTATGGGT GTCTTGSCCC TGCCTGGCTTC
51 GCGTTTTTG GCGGCCATGG TCGTTGCCCG TTGTTGGCGC GTGGCGGCCGT
101 ATGACCGTTT GCGCTGGCG CCGGTGCTGG CGGATGCTTA TAGTCCTCAA
151 CGTTGGCTTGG GCGACGATT TCAAGCAGCG CGAACCTGTC TGTCTGGTGT
201 GCGTTGGGGT GCGCTCTGG CGTGGCTGTC GCGCGGCGT GCGTTCCCG
251 GGGCGGCTTGG GCGCTGGCG CTGCTGATGG TGCCCTTTGTG GATGCCACG
301 TTGGTGGCGG GCGGGCGCGT GCGCTCTGA TTGGGGCGG ACGGCCCTGN
351 GTGGCGCGGG TTGGCAGATA GCGCCATCTC GTTCTGGTAC GCGAAATGTT
401 TTTTNAACCT TCTCTGGTTC GTCAGGGGG CATATCAGGG TTGTTGGCAA
451 GTGCGCGGGG GCGGCCCTCA GCGGCCACNG ACATATGGGGC CGGGGGCGCTG
501 GCGGCCGGTTT TGCGGACATG AAATGCCCGT TTGCGGCCGG TGGCTTGGCG
551 GCGGCCGTCTT CCTGTTCTTC CTGTAATGTT TTTCGGGGTTT CGGGCTTGGCA
601 TTGCTGCTGG GCGGCCACCG TTATGCCAG GCGGAAGCTT AAAATTACCA
651 TTGGTGTCTGG TTGCGAACCTG ATATGCCCGT TGCTTCCGGT CTNGTGTGGC
701 TTGGTGTGGG GTAAACNGCG GCGCGGAGGGT TGCTGTATGC GTGGTGGCGC
751 AGGGCGGGGG TTTCGGATAA GGCGNCTTCC CCTCTGATGC CGTCGCCGCC
801 GCAGCTGGTC GGGGAATATG TGCTNCTGG TTGTCGCCGG CGGGTGTNGT
851 CTGTTGCTGG CCTGTTCTC TTGTTGGCAA TTGTTGTGAA AGCGTGTGGC
901 GCGCGGGAAT CTGGCGCTCT TGTAATGGAA AGTGAACACTT GGCAGGGCGGT
951 GTGGAAACTT NTGCOCTCTT CGGGCGCGG GGTGTATGGC GCGGCCGTTT
1001 TTGGTGTGGT GTATGGCGGG CGCGCGCGGG GGTGTGGGGG GATGCUGCGGG
1051 CTGATGTTT TCGCGCTTGT TGTTGGCGG TTGTTGTGTT CGGGCGGCGT
1101 GCTGCTGCTT NATCCGGAATG GAGCGCGCTTC GTTGGCGCTGG CTGCTGGCGA
1151 TGTATGCCCT GCTGGCGTAT CGCTTGTGG CAAAGATGT TTATCAGCC
1201 TGNGATGCGC TCGCCGGCGGA CGACGOCAGG CGGGCGGGGG GTTGGGGTGC
1251 AAACGGCTT CAAGCACGCT GCGCGCATAC GTTCCCCTC TTAAACCGG
1301 CGTTGGGGCG CGTGTGACT TTGGCGGGG CAACCTGCGT GGGCGNATT
1351 GCGGCCAACCT TTGTCNTGCT CGCTNCNCAG TGCGAGACCC TGACGACTT

```

1401 GATTATGCC TATNTGGGC GCQGQGGTGA NGATAATTAC GCGCGGGCGA
 1451 TGGCTGAC ATTGCTGTG CGGGCGTTG CGCTGGGTAT TTTCCTGCTG
 1501 TTGGACCGGG GCGAAGCGGG AAAACGGACG GAAACGTTAT AA

This encodes a protein having amino acid sequence <SEQ ID 578>:

5 1 MDGGRWAVWG AFALLPSAFI AAMVVAAPLWA VAAVDGLAWA AVLSDAYMLK
 51 RLAWTVEFQAA ATCVLVLPLG VEVAVWLARL AFPGRALVLA LLMLPFVMP
 10 101 LVAGVGVLRL FGADGLXWRG WQDTPYLLLY GNFFFXLPVL VRAAYOGFVQ
 15 151 VEAARLQTAX TLGAGAMWRRR MDIEMPVLRP WLAGGVCLVF LYCFSGFGGLA
 20 201 LLLGGSRYAT VEVEIYQLVM FELDMAVASV LVWLVXGVTA AAGILYAWFG
 25 251 RRAVSDKAVS PVMPSPQSV GHEYVLLAFIA AVXSVCCLFX LLAIVVKAWS
 30 301 AGESWRVLMSE SETWQAQWNT XRFSAVVYAA ANULCVVYAA ARRSAAWMRG
 35 351 LMFLPFMVPSP VCVSAGVLLL XFQWTASLPL LLAMYALLYA PIVAKDVLSA
 40 401 XDALPPDYGR AAAGLNGANGF QTACRITFPF LKPAIIRRGLT LAAATCVGEF
 45 451 AATLFKSRXE WQTLTLLIYA YXGRAGKDXNY ARAMVLTLLL ARFALGXKLL
 50 501 LDGEGGGKRT ETL*

ORF139a and ORF139-1 show 96.5% homology over a 514aa overlap:

orf139a.pep	MDGGRWAVWGAFALLPSAFI AAMVVAAPLWA VAAVDGLAWA AVLSDAYMLKRLAWTVFQAA
orf139-1	MDGGRWVVGAFALLPSAFI AALAVMVWAPLWA VAAVDGLAWA AVLSDAYMLKRLAWTVFQAA
orf139a.pep	ATCVLVLPLGPVVAWVLARLAFPGRALVRLIMLPFVMPVTLVAGVGVIAALPFGADGLXWRG
orf139-1	ATCVLVLPLGPVVAWVLARLAFPGRALVRLIMLPFVMPVTLVAGVGVIAALPFGADGLXWRG
orf139a.pep	WQDTPYLLLYGNVFXLPVLVRAAYQGFVQVPAARLQTAXTLGAGAMRFRWDIEMPVLRP
orf139-1	RQDTPYLLLYGNVFXLPVLVRAAYQGFVQVPAARLQTARTLGAGAMRFRWDIEMPVLRP
orf139a.pep	WLAGGVCLVFLYCFSGFLGALLLGGGSRYATVEVEIYQLVMFELDMAVASV LVLWLVXGVTA
orf139-1	WLAGGVCLVFLYCFSGFLGALLLGGGSRYATVEVEIYQLVMFELDMAVASV LVLWLVXGVTA
orf139a.pep	AAGLLYAWFGRRAVSDKA VSPVMPSPQSVGEYVLLAFAAAAVXSVCCLFXLLAIVVKAWS
orf139-1	AAGLLYAWFGRRAVSDKA VSPVMPSPQSVGEYVLLAFAAAAVLSVCCLFPLLAIVVKAWS
orf139a.pep	AGESWRVLMSE SETWQAQWNTXRFSAAVYAAAVLGVYAAAARRSAWRGLMFLPEMVSP
orf139-1	AGESWRVLMSE SETWQAQWNTLRFSAAVYAAAVLGVYAAAARRSAWRGLMFLPEMVSP
orf139a.pep	VCVSAGVLLLPQWTASLPLLLANALLYA PIVAKDVLSAKDXDALPDDYGRAAAGLGANGF
orf139-1	VCVSAGVLLLPQWTASLPLLLANALLYA PIVAKDVLSAWDALPDDYGRAAAGLGANGF
orf139a.pep	QTACRITFPPLLKPALRRLGLTAAATCVGFEAATLFXSRXEWQTLTLLIYAYXGRAGKDXNY
orf139-1	QTACRITFPPLLKPALRRLGLTAAATCVGFEAATLFLSRPWEQTLTLLIYAYLGRAGEDNY
orf139a.pep	ARAMVLTLLAFAFGXFLLLDGEGGGKRRTETLX
orf139-1	ARAMVLTLLAFAFGIFFLLDGEGGGKQTETLX

Homology with a predicted ORF from *N.gonorrhoeae*

ORF139 shows 95.2% identity over a 189aa overlap with a predicted ORF (ORF139ng) from

55 *N.gonorrhoeae*:

orf139.pep	AGESWRVLMSE SETWQAQWNTLRFSAAA	30
orf139ng	QSVGEVYLLAFSVAVLSVCCLFPLSAIVVKAWSAGE SRRVLMSE SETWQAQWNTLRFSAAA	327
orf139.pep	YAAAALGVVYAAAPARRSAWRGLMFXPFMVS PCVSAGVLLLYPQWTASLPLLLAMYL	90
orf139ng	VFAAAALGVVYAAAPARRLWWRMLGLVFLPFMVS PCVSAGVLLLYPGWTASLPLLLAMYL	387

-332-

5	orf139.pep	LAYPIVAKDVLISAWDALPPDYGRAAAGLGGANGFQTACRITFPPLKPALRRGLTLAAATCV 	150
	orf139.ng	LAYPFIVKADVLISAWDALPPDYGRAAAGLGGANGFQTACRITFPPLKPALRRGLTLAAATCV 	447
	orf139.pep	GEFATLFLSRPEWOTLTLLIYAYLGRAGEDNYARAML 	189
	orf139.ng	GEFATLFLSRPEWOTLTLLIYAYLGRAGEDNYARAMLVTTTLLSFAFVCIFFLLDNGEGG 	507

The complete length ORF139ng nucleotide sequence <SEQ ID 579> is predicted to encode a protein having amino acid sequence <SEQ ID 580>:

10	1 MDGCRWAVRG AFSLLPSAEL AVMMVAPLWA VAYDGLAWR AVLSDAYMLK 5 RLAWTVFQAA ATCCLVPLPG VPVANVLARL AFGPRALVIA LMLPFVMTM 101 LVAGVGLWL FGADGLWRG RQDPYPLLYN GNFVFNLPVN VRAAYQGFAQ 151 VPAARLQAT TLGAGAWPRF WDIEMPVILP RLWAGGCVLVA TCGXFGFLGA 201 LLNGGSRYAT VEVEIYQLVM FELDNAGASA LWLVLGVGTA ARGLLYAWFGT 251 RRASDVDAKS PVMSPSPQSVS GEYVLLFASV AVLSVCLFP LAUWKVAMS 301 AGESRVLME SETQWQAVNT LFSAASAAVF AVLVGVVYAA ARRLLWVMIG 351 LVLPEMFVSP VCUSCAGVLL YPGWTASLP LLAMYALPDR FVPAFVDSLW 401 WDALPDYDGR AAAALGANFG QATCRATFLPK LKPALRGLT LAAATVCGFET 451 AATLFSLRPE QWJLTTLIYA YLGРАGEDNY ARAMVLTLLL SAVFCVCFILL 501 LONGEGKRRT E*#
15	
20	

Further work revealed a variant gonococcal DNA sequence <SEQ ID 581>;

This corresponds to the amino acid sequence <SEQ ID 582: ORF139ng-1>:

55 1 MDGRCWAVRG AFSLLPSAFL AVMVVAPLWA VAAVDGLAWR AVLSDAYMLK
51 RLAWTIVFOA ATCVLWPLGP VPVAVLWLRL APPGRALVLB LMLWLFQMFAD
101 LAVGVLGVL FGADDLWRG RODPTTYLLN GNFVNFLVPL VRAYAQFQMFAD
151 VPAARLQAT TLGAGAWLWM WDINPVLPR WLACGGCLVF CGTSQGFLF
201 LLNGGSRYAT VEVEIYIWLRM FELONAMASA LWWLVLGVTB AGULLYAWFG
251 RRASDVDAKS PVMSPSPQSV GEYVLLAFSV AVLSVCLCLFP LSATVAKM
301 AGESRVLME SETQWAVNT LRFSAAAFAV LAVLGVYVA ARRLLWVAMG
351 LVLPEFMVSP VCVSAGVLL YPGWTASLPL LLANYALLAT PVFKVDSLVA
401 WDALPDYDGR AAALGANGF QTACRITFLP LKPALRGLRT LAATCVGEOF
451 AATLFLRSP EWTQTLTLLYA YLGRAGEDNY ARAMVLTILL SAFVCIFLL
501 LDNGGGKRT ETL*

60

ORF139ng-1 and ORF139-1 show 95.9% identity over 513aa overlap:

	orf139ng	MDCRCWAVRGAESIILPSAFLAVMVVAAPLWAVAAYDGLAWRAVLSDAYMLKRLAWTVFQAA : : : : : : : : : : : : MDGRRWVUVNGAFALLPSAFLAVMVVAAPLWAVAAYDGLAWRAVLSDAYMLKRLAWTVFQAA
5	orf139-1	MDGRRWVUVNGAFALLPSAFLAVMVVAAPLWAVAAYDGLAWRAVLSDAYMLKRLAWTVFQAA
	orf139ng	ATCVLVLPLGVFWAVULARLAFFGRALVRLRLMLPFVMPMLVAGVGVLALFGADGLLWRG : : : : : : : : : : : ATCVLVLPLGVFWAVULARLAFFGRALVRLRLMLPFVMPMLVAGVGVLALFGADGLLWRG
10	orf139-1	ATCVLVLPLGVFWAVULARLAFFGRALVRLRLMLPFVMPMLVAGVGVLALFGADGLLWRG
	orf139ng	RQDTPYLLLYGNVFFNLPLVLRRAAYGFAQVPAARLQTARTLGAGAWRRFDIEMPVLRP : : : : : : : : : : : RQDTPYLLLYGNVFFNLPLVLRRAAYGFAQVPAARLQTARTLGAGAWRRFDIEMPVLRP
15	orf139-1	WLACCVCLVFLYCFSGFGLALLLGSGRYATVEVEIYQLVMFELDMAGASALVWLVLVGTAA : : : : : : : : : : WLAGGVCLVFLYCFSGFGLALLLGSGRYATVEVEIYQLVMFELDMAVASVLCVLLVGTAA
	orf139ng	AAGLLYAWFGRRAVSDKASeVFMSSPPQSIVGEYVLLAFSAVLSVCCFLPFLSAIVVKAWS : : : : : : : : : AAGLLYAWFGRRAVSDKASeVPMSPQSIVGEYVLLAFSAVLSVCCFLPFLSAIVVKAWS
20	orf139-1	AAGLLYAWFGRRAVSDKASeVPMSPQSIVGEYVLLAFSAVLSVCCFLPFLSAIVVKAWS : : : : : : : : AEGESRVRVLMESETQAVWNTRFSAAAVERAAVLGVVYAAARRLIVNMRGLVLFPTMVS
	orf139ng	AEGESRVRVLMESETQAVWNTRFSAAAVERAAVLGVVYAAARRLIVNMRGLVLFPTMVS
25	orf139	AEGESRVRVLMESETQAVWNTRFSAAAVERAAVLGVVYAAARRSAAVNMRGLMFLPFRMVS : : : : : : : AEGESRVRVLMESETQAVWNTRFSAAAVERAAVLGVVYAAARRSAAVNMRGLMFLPFRMVS
	orf139ng	VCVSAGVLLLYFGWTASLPLLLAMYALLAYFPVAKDVLISAWDALPPDYGRAAAGLGANF : : : : : VCVSAGVLLLYFGWTASLPLLLAMYALLAYFPVAKDVLISAWDALPPDYGRAAAGLGANF
30	orf139-1	OTACRITFPPLLKPALRRGTLTAAATCVGFEFAATLFLSRPEWQTLLTLYAYLGRAGEDNY : : : : : OTACRITFPPLLKPALRRGTLTAAATCVGFEFAATLFLSRPEWQTLLTLYAYLGRAGEDNY
	orf139ng	ARAMVLTLLSAFAVCIFLLLDNGEGGKRTETL : : : ARAMVLTLLSAFAVGIFLLLDGGEGGKQTETL
35	orf139-1	ARAMVLTLLSAFAVGIFLLLDGGEGGKQTETL

Based on the presence of a predicted binding-protein-dependent transport systems inner membrane component signature (underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 70

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 583>:

1	ATGGACGGCT	GGACACAGAC	GCTGTCCGGC	CAAACCCCTGT	TGGGCATTTC
51	GGCGCGCGCA	ATCATCTCTCA	TTCTGATTTT	AATCTGCTAGA	TTCCGCGATCC
101	ACGGCGCTGT	GACACTGGTC	ATCGTCAGGC	TGGTACCGGC	TTTGGCAACC
151	GTTTGCCCA	CAGGGCAAGCAT	TGTCAAAGAC	ATACTGGTCA	AAAACCTTCGG
201	CGGCAAGCTC	GGGGGGGTGG	CGCTTCTGGT	CGGCGCTGGGC	GGCGATGCTCG
251	AACTGTTGGT	C...			

This corresponds to the amino acid sequence <SEQ ID 584; ORF140>:

50	1	MDGWQTLSA	QTLIGISAAA	IIILILILIVR	FRIHALLTLV	IVSLLTA LAT
	51	GLPTGSIVKD	ILVKNEFGTL	GGVALVGLG	AMLERLV..	

Further work revealed the complete nucleotide sequence <SEQ ID 585>:

55	1	ATGGACGGCT	GGACACAGAC	GCTGTCCGGC	CAAACCCCTGT	TGGGCATTTC
	51	GGCGCGCGCA	ATCATCTCTCA	TTCTGATTTT	AATCTGCTAGA	TTCCGCGATCC
	101	ACGGCGCTGT	GACACTGGTC	ATCGTCAGGC	TGGTACCGGC	TTTGGCAACC
	151	GGTTTGCCCA	CAGGGCAAGCAT	ATACTGGTCA	AAAACCTTCGG	

201 CGGCACGCTC GGCAGGGCTGG CGCTTCTGGT CGGCCTGGC GGGATGCTOG
 251 GACSTTTGGT CGAAACATCC GGCGGGGACAG AGTCGCTGGC GGAGCGCTG
 301 ATCCGGATGT TCGGGGAAAA ACGGGACCGG TTGGGGCTGG GGGTGGCTC
 351 GCTGATTTC GGCTTCCCGA TTTCTTCTGA TGCCGGACTA ATCGTCATGC
 5 401 TGCCCATCGT TGCTGCCAAC CGAACGGCGA TGAAACAGGA CTACTGCC
 451 TTCCCGCTTG CCTCCATCGG CGCATTTTCC GTCAATGACCO TTCTCTGGC
 501 GCCCCATCGG GGCGCGATTC CGCGTTCGGA ATTTCAGGC GCGACATCG
 551 GCGAAGTTTG GATTTGGGGT CTGGCGACCG CCTTCATCAC ATGGTATTTC
 10 601 AGCGGCTATA TGCTGGCAA AGTGTGGGG CGCACCATCC ATGGTCCCCT
 651 TCCCGAACCTG CTCAAGGGCC CGAACGAAAGA CAACGACCTG CGGAAAGAAC
 701 TGCGGAAAGC AGGAACGCTC GTCCGCGATCA TGCTGATTC CATGCTGCTG
 751 ATTTCCTCTG ATACCGCGGT ATCGGCGCTC ATACGGGAAA AACCTGTAAG
 801 TCGCGCGAAGA ACCTGGGCGA AGAACGAAATAATCGCT TGCGACCGA
 851 TCGCGCTTGT GATTTCTGA TTGGTGGCAG TGTTTGTCTT GGAGCGAAAA
 901 CGCGGGAAA CGCGCGGCG GTTGGAAAA ACCGGTGGACG GGGCACTCGC
 951 CCCGGCTGT TCGCGTGGT TCAGTACCGG CGCGGGCGGT ATGGTGGCG
 1001 CGCTTTGGC CGCTTCCGGC ATCGCGAACGG CACTCGCGA CAGCATGGCG
 1051 GATTGGCGA TTCCGCGCT TTGGGGCTGT TTCCCTGTGCG CCTTGGCACT
 1101 GCGTGTGGC CAAGGGTGG CAACCGTCC CGTGACCAACCG CGCGCCCGC
 20 1151 TGATGGCTC TGCGCTGCC CGCCCGCGT TTACCGACTG CGAGCTGCGC
 1201 TGATCGTGT TGCGCAAGGC CGCACGGTCC GTGCGTTGCA GCGACCTTCAA
 1251 CGACTCGCGC TTCTGGCTGC TGCGCGCTGT CTGGACATCGACGATACCGA
 1301 CCAGCGCTGAA AACCTGGACCG TGTCAACAAA CCTCTATCGGC ACTCATCGGC
 1351 TTGGCGCTTG CGCGCATCGT GTGGCCCATC GTCTGA

25 This corresponds to the amino acid sequence <SEQ ID 586; ORF140-1>:

1 MDGWTOTLSAQTLGLISAAA IILILILIVK FRIHALLTIV LVSLLTALATLRT
 51 GLPTGSIWND ILVKNGFTGL GVALLVGLG AMLGLRIVETS GGAOSLADAL
 101 IRMGEKVRP PALGVASLIF GFPFFPDAGL IVMPLIVPEAT ARRMKDVLVLP
 151 FALSIGAIFS VMEVFLPFLPH GPRIIASEPFVG ANIQVULILG LTAFATIWYF
 201 SGYMLGRVNL RTIHVVPFL LSGGGTQNDL PKEPAKAGTV VAIMLIPML
 251 IFLNTGVSLA ISKELVSAADS TWVTAKIIG STHIALLISV LVALFVLGRK
 301 RGESEGALEK TVDGALAPVC SIVILITGAGG MFPGVLRASC IGKALADMSA
 351 DLGIVLVLGC FLVALARIA QGSATVALTT AAALMAPAVA AAGETDWQLA
 401 CIVILATAQS VGCSHFNDSG FWLGVRLDLM DPTTILRKTWT VNQTLIALIG
 451 FALSELLFAI V*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF140 shows 95.4% identity over a 87aa overlap with an ORF (ORF140a) from strain A of *N.*

meningitidis:

40		10	20	30	40	50	60
	orf140.pep	MDGWTOTLSAQTLGLISAAA IILILILIVK FRIHALLTIV LVSLLTALATLRT					
	orf140a	MDGWTOTLSAQTLGLISAAA IILILILIVK FRIHALLTIV LVSLLTALATLRT					
45		10	20	30	40	50	60
	orf140.pep	ILVKNFGGTLLGVALLVGLGMLERLV	70	80			
	orf140a	MDGWTOTLSAQTLGLISAAA IILILILIVK FRIHALLTIV LVSLLTALATLRT					
50		10	20	30	40	50	60
	orf140.pep	ILVKNFGGTLLGVALLVGLGMLERLV	70	80	90	100	110
	orf140a	MDGWTOTLSAQTLGLISAAA IILILILIVK FRIHALLTIV LVSLLTALATLRT					
55		70	80	90	100	110	120

The complete length ORF140a nucleotide sequence <SEQ ID 587> is:

1 ATGACGGCT GGACACAGAC GCTGTCGGC CAAACCTGT TGGGATTT
 51 GGCGCGGCA ATCATCTCA TTCTGATTTC AATCGTCAAA TTCCGCGATCC
 101 ACGCGCTGTG GACATGGTC ATCGTCAGCC TGCTGACCGC TTGGCAACC
 151 GOTTCGGCCA CGGGCGGCGAT GTCAACAGC GTACTGGTC AAAACTTCGG
 201 CGCGACGGCTT CGGGCGCTGG CGCTTCTGGT CGGGCTGGCG CGCATGCTGG
 251 GACSTTTGGT CGAACATCC GGCGCGCGAC AGTCGCTGCG GGAGCGCGCTG
 301 ATCCGGATGT TCGGGGAAAA ACGGGACCCG TTGGGGCTGG CGCTTGGCTC
 351 GCTGATTTC GGCTTCCCGA TTTCTTGA TGCCGGACTA ATCGTCATGC
 60 401 TGCCCATCGT GTTCCGACCC CGAACGGCGA TGAAACAGGA CTACTGCC
 451 TTCCGCGCTTG CCTCCATCGG CGCATTTTCC GTCACTGCACG TCTTCTGGC

501 GCCCCATCGG GGCCGGATTG CCGCTTCGGA ATTTTACGGC GCGAACATCG
 551 GCAAGTTTT GATTGGGT CTGGCCACCG CCTTGATCAC ATGGTATTT
 601 AGCGGTATA TGCCTGGCAA AGTTGGGG CGCACCATCC ATGTTCCCGT
 651 TCCGGACTG CTGAGCGCG GCACGCAAGA CAACGACCTG CGGAAAGAAC
 701 CTGCCAACAG AAGAACCGTC GTGCCATCA TGCTGATTC CATGCTGCTG
 751 ATTTCCTCGA ATACCGCGCTG ACAGGCCCTC ATCAGCGAAA AACCTCTAAG
 801 TGCGGACGAA ACCTGGTTG AGACGCGAA AAAATCGGT TGACACCGA
 851 TCGCCCTCT GATTTCGCGA TTGGTGCACAT TTGTTGCTCTT GGAAACCCAAA
 901 CGGGGGAAA GCGGACGGC GTTGGAAAAAC ACCGGTGACG CGGACTCTCG
 951 CCCCGTCTG TCCGGTATTC TGATTACCGG CGCCGGGGGT ATGTTGGGG
 1001 GCGTTTGGC CGCTTCCGGC ATCGCAAGG CACTGGCGA CAGCA'GGG
 1051 GATTGGCTCA TCCCGCTCTT TTGGCTCTGT TTCCCTGTGCG CCTTGGCAGCT
 1101 CGGTGATCCGG CAAAGTTCGG CAACCTTCGG CCTGACCAACC CGCCCGGGCG
 1151 TGATGGCTCG TCCGGTGC GCGCCGGGT TTACCGGACTG GCAGCTCGCC
 1201 TGATGCGTAT TGCGAACGGC GGCAAGTTGC GTCGGTTGCA GCACACTCAA
 1251 CGACTCCGGC TTCTGGCTGG TCCGGCGCTT TTGGACATG GACGTACCGA
 1301 CCACGCTGAA AACCTGGACG GTCAACCAAAC CCCTCATCGC ACTCATCGGC
 1351 TTGGCTCTGT CGGACGCTG GTTCGCCATC GTCTGA

This encodes a protein having amino acid sequence <SEQ ID 588>:

20 1 MDGWTQTLSA QTLGLISAAA IILILILIVL FRIHALLLTIV IVSLLTALATL
 51 GLPTGSIVND VLVRKNSFGT GLVALLVGLG AMLGRIVLVE TS GGAA\$SLADAL
 101 IRMFGEKRAF DALFVGASLIF GFPPIFFDAGL IVMPLIVFAT ARRMKDVL
 151 FALASIGAFS VMHVFPLPHP GFPIASEFYG ANIGQVLLIG LPTAFITWYF
 201 SGYMLCKVLG RTIHPVPCGEL LSQQTDNDL PKEPKAGTV VAIMLPMLL
 251 IFLNTGVSA LISEKLVSADAE TWQTAKIIIG STPIEALLISV LVALFVLGRK
 301 RGEESGSALEK TVDGALAPV SLLITGAGG MEGGVILRASG JKGALADMSA
 351 DLSIPVLLGC FLVALALARIA QGSATVALTT AAALMAAPAVA AAGFTOWQLA
 401 CIVLATAAGS VGCSHFNDSG FWLVGRLLDM DVPTTLKTWT VNQTLIALIG
 451 FALSALLPAI V*

30 ORF140a and ORF140-1 show 99.8% identity over a 461aa overlap:

orf140-1.pep	MDGWTQTLSAQTLGLISAAA IILILILIVL FRIHALLLTIV IVSLLTALATLGLPTGSIVND	60
orf140a	MDGWTQTLSAQTLGLISAAA IILILILIVL FRIHALLLTIV IVSLLTALATLGLPTGSIVND	60
35 orf140-1.pep	I LVKNFGTTLGGVALLVGLGAMLGRIVLVE TSGGAQSLADALIRMFGEKRAPFALGVASLIF	120
orf140a	I LVKNFGTTLGGVALLVGLGAMLGRIVLVE TSGGAQSLADALIRMFGEKRAPFALGVASLIF	120
40 orf140-1.pep	GFPPIFFDAGL IVMPLIVFATARRMKQDVLPFALASIGAFSVMHVFLPPHPGPPIASEFYG	180
orf140a	GFPPIFFDAGL IVMPLIVFATARRMKQDVLPFALASIGAFSVMHVFLPPHPGPPIASEFYG	810
45 orf140-1.pep	ANIGQVLLIGLPTAFITWYFSGYMLGKVLRGTRIHVPVELLSGGTQONDLPKEPAKAGTV	240
orf140a	ANIGQVLLIGLPTAFITWYFSGYMLGKVLRGTRIHVPVELLSGGTQONDLPKEPAKAGTV	240
50 orf140-1.pep	V AIMLIPMLLIFLINTGVSALISEKLVSADAE TWQPAKIGSTPIALLISV LVALFVLGRK	300
orf140a	V AIMLIPMLLIFLINTGVSALISEKLVSADAE TWQPAKIGSTPIALLISV LVALFVLGRK	300
55 orf140-1.pep	RGESSGALEK TVDGALAPVCSVILITAGGGMFPGVLRASICGKALADSMADLGIPVLLGC	360
orf140a	RGESSGALEK TVDGALAPVCSVILITAGGGMFPGVLRASICGKALADSMADLGIPVLLGC	360
60 orf140-1.pep	F LVVALALRIAQGSATVALTTAAALMAPAVA AAGFTDWQLACIVLATAAGSVGCHFNDSG	420
orf140a	F LVVALALRIAQGSATVALTTAAALMAPAVA AAGFTDWQLACIVLATAAGSVGCHFNDSG	420
orf140-1.pep	FWLVGRLLDMDVFTTLLKWTWVNOTLILALIGPALSALLFAIV	461
orf140a	FWLVGRLLDMDVFTTLLKWTWVNOTLILALIGPALSALLFAIV	461

Homology with a predicted ORF from *N.gonorrhoeae*

ORF140 shows 92% identity over a 87aa overlap with a predicted ORF (ORF140ng) from

65 *N.gonorrhoeae*:

-336-

The complete length ORF140ng nucleotide sequence <SEQ ID 589> was predicted to encode a protein having amino acid sequence <SEQ ID 590>:

10 1 MDGRTOTLSA QTLGLISAAA IIIILILILIVK FRIRALLTLV IASLLTALAT
51 GLPFTSIVNN VLVKNFGPTG GGVALLVGLS AMGLRVLTS GGAAOSLADAL
101 IRNFGEKFRV FAFGSASFVLF GGFIFPDAGL UMLIPFIVPAS ARRMQDVL
151 FALASVGQAFS VMHVFVPLFGP GPIIAASEYF ANIQVGQLLA TFLATWIFTY
201 SGYMLGKVLG RAIHVPPVPEL LSGCTDSDF PCKEAKAGTV VAVMLIPML
15 251 IFINTGVSAI ISEKLVSADAE TWQOTAKMIG STPVALLSIV LAHLULVGRK
301 RGEGSTLEK TVGDALAPCA SVLITLGAGG MFFGVLBRASG IGKALADSMA
351 DGLIPVFLGC FLVALAIRLA QGSATVALTT ARAALMAPVA AGFTWDQLA
401 CIVLATAGS VGCSHFNDGS EWLWGRSLSDM DVPTLTKWT VNQTLIAFPG
451 FALSALFAI V*

20 Further work revealed a variant gonococcal DNA sequence <SEO ID 591>.

1	ATGGACGCCG	GGACACAGC	GCTGTCGCCG	AAACATTGTG	TGGCGATTTC
51	GGCGCTTCA	ATCATCTCA	TCTTGATTT	AATCGCTAAA	TTCCGATCCC
101	GCGGCTGCT	GACACTGTC	ATGCGCCAGG	TGTCGACGCC	TTGGCAACCC
151	GGTTTGGCC	CAGCGCTGC	CGTCACAGCG	TAATGCTGCG	AAACATTTCG
201	CGCGACGCT	GGCGGCGTGG	CGCTTCTGGT	CGCGCTGGGA	GCATGTGCTG
251	GACGCTTGTG	AGAACACATC	GGCGCCGCAC	AGTCGTTGCG	GGCGCGCTGC
301	ATCCGGATGT	TGGCGAARA	CGACCECACCG	TTCTCCGGT	GGCTGCTGCC
351	GTGTGATTTC	GGCTTCCGAA	TTTCTCTGCA	TGGCGGACTA	ATCTGATCAG
401	TGGCCATCGT	ATTCGGCCAC	GGACGGCGCA	TAACGACAGG	CATCTGCCCC
451	TTGGCGCTG	CCTCCGCTGG	CGCATTTTCCC	TGTCATGACCA	TTCTTCCTCC
501	GCCCCATCC	GGCCCGATTG	CGCCCTTCCGA	TTTATTCAGCG	GGCAATCATCG
551	GCGCAGTTT	GATTTTGGT	CTGCGGCACGG	CTTICATCAT	AGTCGTTATTC
601	AGCGCTTATA	TGCTGCCGCAA	ATGTTGGGG	GGCGCCATCC	ATGTTCCCGT
651	TCCCCGAACTG	CTTCAGCGGC	GAACCGAAAGA	CAGCGGACCCG	GGCGAAAGAC
701	CTGCGAAAGC	AGGAACGGTC	GTGCGCTCTG	TCTGATTCGG	CATCTGCTG
751	ATTTTCCTGA	ATACCGCGG	ATACGGCTTC	ATACGGGAA	ACTCTGTAAG
801	TGGCGACGAA	ACTTGGGTC	AGACGGCAAA	ATATGATCGGT	TGCGACACTG
851	TGCGCTCTCT	GATTTCCTGA	TTGGCCCGAC	TGTTGGCTT	GGACGGCAGAA
901	CGCGGCGAA	GGCGGACAGC	GTGGAAAAAA	ACCGTGGCG	GGCGACTCGG
951	CCCCGGCTGT	TCCGTGATTC	TGATTAACCGG	CGCGGGCGGT	ATGTCGCGG
1001	GGTTTGTGG	GGCTTCCGGC	ATTCGGCAAGG	CACTCGCGCA	CAGCTGCGG
1051	GATTTGGCG	TTCCCGTCT	TTGGGGCTGC	TCTTCTGTC	CTTGGGACT
1101	GGCTGATCCC	CAAGGGTGG	CAACCGTGGC	CTTGACACAC	GGCGCGCCGG
1151	TGATGCTGCC	TGCCGGTGGC	GGCGCGCGTC	TTCAGCTGCG	GCACGTGCGC
1201	TGTGATCTAT	TGGCAAGGGC	GGCAAGGTTG	CTCGGTGCG	GGCCATCTCA
1251	CGACTCGCT	TTCTGGCTGG	TCGCGGCCGT	TCTGGATATG	GACSTTACCGA
1301	CCACGTGAA	AAATCGG	GTCAACAAA	CCCTCATCGC	ATTCATCGGC
1351	TTTGGCTGTG	CGCGACTGCT	GGTTCGATCG	CTGTGA	

This corresponds to the amino acid sequence <SEQ ID 592; ORF140ng-1>:

50 1 MDGRTQLSA QTLLGISAII LILILILIVK FRIRALLTLV IASLLTALAT
51 GLPFTSIVNVLWLRNGKFQGGVALLVGLS AMGLRNLTSVGS QAGSLADL
101 IRMFGKFRAR FAPVGASLIF GFFIFDAGL UMLPPIVFS APPRMQDVL
151 FALASQGKVSF MMVHFLPQHF GPPIAASEYFQ ANIQVGVLG LTPLATWIFTY
201 SGYMLGKVLG RAIHVVPEL LSGQTDSDFE PKEKAKGTV VAVMLIPML
251 IFLNTGVSAL ISEKLVSADA TWQOTAKMIG STPVALLSV LARALVGRK
301 RGESGSTTLEK TVDGALPAC SVLITGAGG MFGSVLRASG IKRALADS
351 DGLIPVFLIGC ELVALALRICA QGSATVALTT ALARIMAPAVA AAGFTWDQLA
401 CIVILATAAGS VGCSHFNDSG FBLVGRLLDM DVPTLTKWT VNQTLIAFIG
451 FALSALLFAI V*

60 ORF140ng-1 and ORF140-1 show 96.3% identity over 461aa overlap:

orf140-1 MDGWTQTLSAQTLGLGISAIIILILILIVKFRIHALLTIVLIVSLLTALATGLPTGSIVND
 orf140ng-1.pep VLVKNPGGTLLGGVALLVGLGAMLGRLVETSGGAQSLADALIRMFGEKRAPFAPGVASLIF
 5 orf140-1 ILVKNPGGTLLGGVALLVGLGAMLGRLVETSGGAQSLADALIRMFGEKRAPFALGVASLIF
 orf140ng-1.pep GFFIPFFDAGLIVMLPIVFATARRMKQDVLPFALASVGAFSVMHVFLLPPHPGPPIAASEFYG
 10 orf140-1 GFFIPFFDAGLIVMLPIVFATARRMKQDVLPFALASIGAFSVMHVFLLPPHPGPPIAASEFYG
 orf140ng-1.pep ANIGQVLLILGLPTAIFTWYFSGYMLGKVGLGRAIHPVPELLSGGTQDSDPKEPAKAGTV
 15 orf140-1 ANIGQVLLILGLPTAIFTWYFSGYMLGKVGLGRAIHPVPELLSGGTQDNDLPKEPAKAGTV
 orf140ng-1.pep VAVMLIPMLLIFLNTGVSALEKLSVADETWQTAKIGSTPIALLISVLVALFVLGRK
 20 orf140-1 VAVMLIPMLLIFLNTGVSALEKLSVADETWQTAKIGSTPIALLISVLVALFVLGRK
 orf140ng-1.pep RGESGSTSTEKTVDGALAPACSVLILITGAGGMFGGVLRASIGKALADSMADLGIPVLLGC
 orf140-1 RGESGSTSAEKTVDGALAPACSVLILITGAGGMFGGVLRASIGKALADSMADLGIPVLLGC
 25 orf140ng-1.pep FLVALALRIAQGSATVALTTAAALMAPAVAAAGFTDWQLACIVLATAAGSVGCSHFNDSG
 orf140-1 FLVALALRIAQGSATVALTTAAALMAPAVAAAGFTDWQLACIVLATAAGSVGCSHFNDSG
 orf140ng-1.pep FWLVRGLLDDMDVFTTLLKTWTVNQTLIAFIGFALSALLFAIV
 orf140-1 FWLVRGLLDDMDVFTTLLKTWTVNQTLIALIGFALSALLFAIV

30 Furthermore, ORF140ng-1 is homologous to an *E.coli* protein:

gi|882633 (U29579) ORF_o454 [*Escherichia coli*] >gi|1789097 (AE000358) o454;
 This 454 aa protein is 34% identical (9 gaps) to 444 residues of an approx. 456 aa
 protein GNTP_BAC1 SW: P46832 [*Escherichia coli*] length = 454
 Score = 210 bits (529), Expect = 1e-53
 35 Identities = 130/384 (33%), Positives = 194/384 (49%), Gaps = 19/384 (4%)
 Query: 88 ETSGGAQSLADALIRMFGEKRAPFAPGVASLIFGFPPIFFDAGLIVMLPIVFATARRMKQD 147
 E SGG+SLA+ R GKR A +A+ P+FFD G I++ PI++ A+ K
 Sbjct: 80 EHSGGAESLNFSRKLGDKRTIAFLTAAFLGKIPVFFFDFGFTIILAPIIYGFAKVAKIS 139
 40 Query: 148 VLPFALASVGAFSVMHVFLLPPHPGPPIAASEFYGAGNQVLLILGLPTAIFTWYFSGYMLGK 207
 L F L + +HV +FPHRGPI+A+ A+IG- I+G+ +I W GY K
 Sbjct: 140 PLKFGPLVAGIMLTVHAWVAPPHGPGPVAAAGLHLADIGWLTIIGIAIS-IPGVVVGYPAK 198
 45 Query: 208 VLGRAIHVPPELL-----SGGTQDSDPKEPAKAGTVVAVMLIPMLLIFLNTGV 257
 ++ + + E-L G T+ SD P A V +++IP+ +I T
 Sbjct: 199 IINKRQYAMSVEVLEQMLQAPASEEGATKLSDKINPPGVA- LVTSLIVIPIAIIMAGT- 255
 50 Query: 258 SALISLSESADETWQTAKIGSTPXXXXXXXXXXXXXXXXXXGRKRGESGSTILEKTVIGALA 317
 +S L+ + + T ++IGS +RG S + AL
 Sbjct: 256 +VSATIMPPSHPLLGTQLIGSPMVALMIALVIAFWLIALRREGWSLQHTSDIMGSALP 312
 Query: 318 PACSVLITGAGGMFGGVLRASIGKALADSMADLGIPVLLGCFVLVALALRIAQGSXXX 377
 A VIL+TGAGG+FG VL SG-GKAL+ + +P+L F+I+LAL+ +QGS
 55 Sbjct: 313 TAAVVILVTGAGGVFGKVLESGVGKALANMLQIMDLPLLPAAIFIISLALRASQGS--AT 370
 Query: 378 XXXXXXXXXXXXXXXXGFTDWQLACIVLATAAGSVGCSHFNDSGFWLVGRLLDMDVPTTLK 437
 G Q + LA G + SH NDSGFW+V + L + V LK
 Sbjct: 371 VAILTTGGLSEAVMGLNP1QCQVVLTLAACFCGGLGASHINDSGFWIVTKYGLSVADGLK 430
 60 Query: 438 TWTWNQTLIAFIGFALSALLFAIV 461
 TWTV T++ F GF ++ ++A++
 Sbjct: 431 TWTWTLLTILGFTGLITWCWVAVI 454

Based on this analysis, including the identification of the presence of a putative leader sequence
 65 (double-underlined) and several putative transmembrane domains (single-underlined) in the

gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 71

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 593>:

```

5      1 ..GATTCGGCA TATCGCCCGT GTATCTTTGG GTTGCGCCCG CGTTCAAACA
51     TTGCTGTGCG CGGTGGCGTC CGGACTCATATA CGATGTGCGA CGCTTTCGAG
101    GCGTATTATTC TGGCGTATACG GGACTGACTT CTCGGGGCTT TGCGGGTTTC
151    AACTTTTGGG CGACAGACCA CGGGCCGAC. GTCGTCCTGA TTTCATCGG
201    CTGATCCGGG CTGATCCAG TGCGCCATTCTT CCTCAACCCCG GTGCGCCGCC
251    CCTTGCCTGG CGCGCGACTG GTGCTGQAGG GTTATTCCTT GGCTCGCCGG
301    CGGTGATATTG CGCGCTCTTT TCCTGTCGGT AGCGGCCTGGA CGCTGATGTC
351    GTTGGCAGGA GCTTATCCGG CAGCATTGTCG CCTGATGCTG CCCTGCCCCG
401    TACTGATGTT TTTCGGTCCG ...

```

This corresponds to the amino acid sequence <SEQ ID 594; ORF141>:

```

15      1 ..DFGISPVYIWW VAAAFKHLLS PWAADSYDVA RFAGVFFAVI GLTSCGFAGF
51     NFGGRHHGRK VVLLIIGLCIG LIVFHFLNP AAAFAAAGAL VLHGYSLARR
101    RVTAASFILLG TGWTIMSLAA YFAAFAFLM LPVLMFPRF ...

```

Further work revealed the complete nucleotide sequence <SEQ ID 595>:

```

20      1 ATGCTGACCT ATACCCCGCC CGATGCCCGC CGGCCGCCAA AAACCCACGA
51     AAAAGCGTGG CGTCTGGCTGT TGATGGCGTT TGCGCTGGTT TGCGCCGGCG
101    TGTTTTCCCA CGATTGGCTGG ATACCTGACG AACCTGGCGT CTAACTACGCC
151    GTCGAAGCAC TGGCAGGCA CGCCACCCCG TTGGCTGCCG ACTCTGTTGG
201    TCAACAGAT TTGGCGATCG CGGGCGCTGTA TTCTTGGGGT GCGCGCGGGT
251    TCAACATTG GTGCTGCGCG TGGCTGCCG ACTCATACGA TGCGCACGC
301    TTGGCAGGGG TATTTTGGCG CGTGTATCGA CTGAGTTCTC GCGCTTTCG
351    CGTTTGCAAC TTTTGGCGA GACACCCAGG GCGCAGGCTG GTCCTGATTC
401    TCTACGGCTG ATATGGCGCT ATTACAGTCTT CCCATTTCTC CAACCCCGCT
451    CGCGCGGCCG TTGGCGCCG CGGACTGTCG CTGACAGGT ATTCTTTGGC
501    TCGCGCGGC GTGATGCGC CCTCTTTCTC CTGCTGGTAGC GGCTGGACGC
551    TGATSTCGTT GCGCAGCAGT TATCGGCAG CATTGGCCCT GATGCTGCC
601    TTGGCGGTAC TAATGTTTT CGCTCGTGG CAAAGCAGGG GTTGTATGTT
651    GAGCGCAGTC GCTCCTACGT CCTTGGCCCT GCCCTTATAG ACCGTTTAC
701    CGCTGCTGTT GCGAACAAAGC CAGCGCGGGC GTTTCGCCGCA ATGGCTCGAC
751    TATCACGTTT CGTGGTACGT CGGGCGCTCG CGGCAAGCTTC AGACGGCATT
801    CAGTTGTTG TTAATATCTGA AAAACCTGCT TTGGTTTGGCA TTGCGCCGGC
851    TGGCGCTGCG GGTGTTGGAG GTTTCGGCGA CGCGCTCTGT TTGCGACGAC
901    TGGGGGATTT GTGGCGCTGT CGNGAGTGTG CGCGTTTTGG TGCTGCTTGC
951    CGCTCAATCGC AGCGTTTGTG AGATAAACCT CGTCGGCTG CTTCGGCCG
1001   TTGGCCCTGG CGCGCGGGGG CAACCTGACA GCCTGAGGG CGCGCGGGCG
1051   CGCTTGTCTA ACTGTTCTC CATTATGGCG CCTCGAGCTGT TTCCGGTGT
1101   CCTGGTGACG GCGTTTTCTG CCTATGAAATA CGCGTGGCCG GCGAAGCTTG
1151   CGGACAGCGGC CGCTCTGCG CGCGCTGTT ATTGTCTCTGA TATCGATGCC
1201   ATTCCCGATGG CGGTGGCGT ACTGTTCACAC CCTCTGGCG TTGGGGCAT
1251   TACCCGAAAA AACATACGGG CGACGGCGGG GTTACCAAC GGGCATCGCT
1301   CGCTTACCTT GACCTGGCGT TTCTGTGAGA CGCTGTTCTC GCGCTGGCTG
1351   GACCGGGCGA AAAGCCACG CGCGGTGCTC CGGAGTATGG AGGCATCGCT
1401   TTCCCCGGAA TTGAAAGGG AGCTTTCAAGA CGCGCATGAG TTGATCGCA
1451   TAGGGCGGGC CGACCTTCGAC AGCGCGATG TTGGAGCGCA GTACGGCAC
1501   TTGCGCCACG CGGTGGCGCA TGTAACATGG CGCTACCGCA TCTCTCTCT
1551   GCGCCAAAAT CGGGATGCGC CGCAAGGCTG CGAGACGGTT TGCGAGGGTG
1601   CGGGTCCGGG CAACAAAGAC AGTAAGTTCG CACTGATAAG GAAAATCGGG
1651   GAAAATATATAA

```

This corresponds to the amino acid sequence <SEQ ID 596; ORF141-1>:

```

55      1 MLTYTFPDAR PFAKTHEKPW LLLLMAFAWL WFGVFSHDLW NPDEPAVYTA
51     VEALAGSPTF VLAHFGQTD FGIPFVYLNW AAAFKHLLSP WAADSYDAAR
101    FAGVGVFAVG LTSGCFAGFN FGIRHHGRS VVLLIIGCLG IVPFHFLNP
151    AAAFAAAGAL VLHGYSLARRR VIAASFLGTT GTWTLMSLAA YFAAFAFLM
201    LPVLMFPRFW QSRRMLTAV ASIALAFPLM TYVFLLAKT QPALFAQWL

```

251 YHVVEGTFFGVY RHHVQTAFLSLF YKLKNLWWPA LPLALPLAWT QVRFLSTD
 301 WGGILGVVWML ALVLLVLAAPNQ QRFQDNLWLW DPLPLAFGAA QLDSLRRGAA
 351 AFWNNFGMIA FGLEFWLMT GFEMANFYGGK AKLAERAAYF SPYVPMFDIP
 401 IIMAVWALFT EPLKWLAITR NIKRGQAVTN LMLTFLPWL
 451 DAARKSHAPVW RSMEASLSPK LKRELSDC CIGIGGDLDE TRIWVQYGT
 501 LPHRVRGDVQC RYRIVLPLLPOA ADAPQGKQTVW QKGARPRNKA SKFALIRKIG
 551 ENI*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

- 10 ORF141 shows 95.0% identity over a 140aa overlap with an ORF (ORF141a) from strain A of *N. meningitidis*:

				10	20	30
	orf141.pep			DFGIFSPVLYNAAFKHLLSPWAADSYDVA		
15	orf141a	WNDEPAPAVYTAVEALAGSPTPLVAHLFGQIDFGIPFPVLYNAAFKHLLSPWAADSYDVA		:: : : :		
		40 50 60 70 80 90				
			40 50 60 70 80 90			
20	orf141.pep	RFAGVFFFAVIGLTS CGFAGFNFLGRHHGXXVVLLILIGCIGLPIVHNPAAAAAFAAAGL		: : : : :		
	orf141a	RFAGVFFFAVIGLTS CGFAGFNFLGRHHGXXVVLLILIGCIGLPIVHNPAAAAAFAAAGL				
		100 110 120 130 140 150				
25	orf141.pep	V LHGYSLARRR V I AASFLCTGTW T I M S L A A Y P A A F A I M L P L P V L M F F R P		100 110 120 130 140		
	orf141a	V LHGYSLARRR V I AASFLCTGTW T I M S L A A Y P A A F A I M L P L P V L M F F R P Q S K R R L M L T A				
		160 170 180 190 200 210				
30	orf141a	VASLAFAFLPLMTVYPLLAKTQPALFAQWLDDHVFGTFGGVRHIQTAFSLFYLNKLNLWF		220 230 240 250 260 270		

The complete length ORF141a nucleotide sequence <SEQ ID 597> is:

	1	ATGGTCGACTT	ATACCGCCCG	TGATGCCGCG	CCGGCCGCCA	AAACCCACGCA
35	51	AAAGCCGCTT	CTGTTCTGTG	TGAGTGCGTT	TGGCTCCTGG	GGCCGGCGGC
	101	TGTTTTCCCA	CGATTGTTGG	ATCTCTGACG	AACCTGCGCT	CTATACCGCC
	151	GTGCGAACAC	TGGCGACGAG	CCCCACCCC	TTGGTTGGCC	ATCTGTTCTGG
	201	TCAAATGAT	CGCCGCGATC	CGCCGGTGTG	CTTTGGGGT	GGCCGGCGGT
	251	TGAAACATCT	GCTGCGCCG	TGGCTGCGC	ACCGCTATGAT	TGGCGCACGC
40	301	TTGGCCGGG	TGTTTTCGG	CCTGGTGGGA	TCGACTCTCC	GGCGGTTGG
	351	CGGTGTTCA	TTTTGGCG	GACGACACCG	GCGCCAGCGTC	TGCTCTGATTG
	401	TCATCGCT	TTAGGGCGT	ATTCAGACCG	TAACACTTCT	AAACCCCGCT
	451	GGCGCCGCT	TTGGCGCG	CGGACTGTCG	CTGCGGTTCT	ATTCCTTGGC
	501	TGGCGGCG	GTGATGGCG	CTCTTTCTCT	TCGCGGTACG	TGTTGGACGC
45	551	TGATGTTGCT	GCGACGACT	ATCTGGCGGC	CATTGGTCCC	GTATGCTGCC
	601	CTGGCGCTTC	TGTTGTTTCT	CGCTGGCTGG	AAACAGCAGG	GTTCATGATGT
	651	GACGCCAGTC	GGCTGGCTTC	CGGCGGCTCT	GGCGCTTATG	ACCGTTTACCG
	701	CGCTGCTCT	GGCAAAATCG	CGACGGCGGC	TGGTCCGCCA	ATGGCTGCGAC
	751	GATCAGCTT	TGGCTGACTT	CGGGCGGGTC	CGCCGACATTTC	AAAGCAGCATT
50	801	CAGTTGTTT	TAATCTGCT	AAACACTGCT	TGTTGTTGCA	TTGGCTCGGC
	851	TGGCGCTGGC	GGTTTGGACG	TTTGTCGCGA	CGGGCTCTGT	TTGCGACGAC
	901	TGGGGGATT	TGGGGCGCTG	TCGATGCTCTT	GCGCTTGTG	TGCTCTTGTG
	951	CGTCATACTG	CAAGCTTTTC	AGGAAATTAAC	CTGTCGGCTC	CTTCGGCGGC
55	1001	TGGCTCTT	CGGGCGGGCG	CAACTGGACA	GCTCTGAGAC	GGGGCGGGCG
	1051	GGCTGTTGCA	ACTGGTTCGG	CATTATGGGG	TTGGGACTGT	TTGGCTGTTG
	1101	CTGTGTTGAGC	GGCTTTTTCG	CCATGATTA	CGCTGSSGCC	CGGAAGCTG
	1151	CGGAAGCGC	CGCTCATTTTC	AGCCGGTATT	ATGTTCTGCA	TATCGATCC
	1201	ATTCGGATGG	GGCTTGGCGT	ATGCTTCACA	CCCTTGTGGC	TGTGGGGCGAT
	1251	TACCGGCCAA	AAACATACCGC	CGAGCGACGG	GGTTAACCAAC	GGGGGGCGAG
	1301	GGGTACCTT	GGACCTGGCT	TGTCGATGA	CGCTGTTCTC	GGGGTGTGCT
60	1351	GACGGCGGAA	AAAGCAGCG	CGGGCTCGTC	CGGAGATGAT	AGGCATCGCT
	1401	TTCCGGCGGA	TTAAAACCGG	AGCTTCAGA	CGGCATGCGA	TGATATGACA
	1451	TAGGGCCGGC	CGACCTACAA	CGGGCGATG	TTGGACGCCA	GTACGGCCACA
	1501	TGGCGGAAAC	GGCGTGGCG	TGTCGATACG	CGCTGACGCC	TGTCGCTCTT
	1551	GGCCCCAAAC	GGCGTGGCG	CGACCGCTG	CGACAGGTT	TGGCAGGGTG

1601 CGCGCCCGCG CAACAAAGAC AGTAAGTTCG CACTGATAAC GAAAACGGG
 1651 GAAAATATAAT TAAACACAC AGATTGA

This encodes a protein having amino acid sequence <SEQ ID 598>:

5 1 MLTYTTPDAR PPAKTHEKPW LLLLMAFPALW WPGVFSHDLM NFDPEAVYTA
 51 VEALAGSPTP LV AHLFGQOID FGIPVPLWV AAAFKHLLSP WAADPYDAAR
 100 101 FAGVFFAVVU LTSCGFAGFN FLGRHHGRSV VLLILGCIGL IITVHFLNPA
 151 151 AAAFPAAGGLV LHGSYSLARR VIAASFLGT GTWILMSLAAA YEEAFLMLP
 201 201 LPVLMPFRPW QSRRLMIAAV ASLAFAPIIM TVYPLLIAKT QPALEAQWLD
 251 251 DHVSGFTRGGV RHIQTAFLSF YLKLKNLWEA LPALPLAVWT VCRTRLFSTD
 301 301 WGILGVVWMML AVLVLAVN P QRFDNLVWL LPPALFLGAA QLDSLRLRKGAA
 351 351 AFVNWFGIMA EGGLEVELWTF GFEMMYGWNP AAKLAERAAYF SYYVVPDIDP
 401 401 IPMRAVLF PTLWLMWAITRK NIRGRQAVTN WARGVLTITWA LMLTFLPWL
 451 451 DAAKSHAPVW RSEMEASLSPE IKRELSDGIE CIDIGGGDLH TRIVWTQYGT
 501 501 LPHRVGDVQC RYRIVRLPQN ADAEQLGWQTV WQGARPRNKD SKFALIRKTG
 551 551 ENIILKTTD*

ORF141a and ORF141-1 show 98.2% identity in 553 aa overlap:

orfl41a.pep	MLTYTTPDAR PPAKTHEKPW LLLLMAFPALW WPGVFSHDLM NFDPEAVYTA
orfl41-1	MLTYTTPDAR PPAKTHEKPW LLLLMAFPALW WPGVFSHDLM NFDPEAVYTA
orfl41a.pep	LVAHLFGQIDDGIPPVYLWVAAAFAKHLSSPWAADPYDAAR FAGVFFAVVGLTS ^G FAGFN : : : : : : :
orfl41-1	LVAHLFGQTDPGIPPVYLWVAAAFAKHLSSPWAADSYDAAR FAGVFFAVVGLTS ^G FAGFN : : : : : : :
orfl41a.pep	FLGRHHGRSVVLLILGCIGL IITVHFLNPAAAAFAAGGLV LHGSYSLARR VIAASFLGT : : : : : : :
orfl41-1	FLGRHHGRSVVLLILGCIGL IITVHFLNPAAAAFAAGGLV LHGSYSLARR VIAASFLGT : : : : : : :
orfl41a.pep	GWTLMSLAAAYPAAFALMPLPVLPMFWRPQSRLMLTAVASLAFA PLMLTVYPLLIAKT : : : : : : :
orfl41-1	GWTLMSLAAAYPAAFALMPLPVLPMFWRPQSRLMLTAVASLAFA PLMLTVYPLLIAKT : : : : : : :
orfl41a.pep	QPALFAQWLDHVFGTFGGGVRHQTAFLSFYLYLKNNLWFA PLPALPLAVWTVCRTL ^R FSTD : : : : : : :
orfl41-1	QPALFAQWLDHVFGTFGGGVRHQTAFLSFYLYLKNNLWFA PLPALPLAVWTVCRTL ^R FSTD : : : : : : :
orfl41a.pep	WGILGVVWMMLAVLVLLAVNPQRFDNLVWL LPLPPLALFGAA QLDSLRLRGAA AFVNWF ^G IMA : : : : : : :
orfl41-1	WGILGVVWMMLAVLVLLAVNPQRFDNLVWL LPLPPLALFGAA QLDSLRLRGAA AFVNWF ^G IMA : : : : : : :
orfl41a.pep	EGLPEAVFLWTFGFPMNYGWPAKLAERAAYFSPYVWDIDP ^I PMAVAVLFTPLWLMWAITRK : : : : : : :
orfl41-1	EGLPEAVFLWTFGFPMNYGWPAKLAERAAYFSPYVWDIDP ^I PMAVAVLFTPLWLMWAITRK : : : : : : :
orfl41a.pep	NIRGRQAVTNWAAGVTLTWALLMTLFLPWLDAAKS ^H A ^V VRSMEASLSPELKRELS ^D GIE : : : : : : :
orfl41-1	NIRGRQAVTNWAAGVTLTWALLMTLFLPWLDAAKS ^H A ^V VRSMEASLSPELKRELS ^D GIE : : : : : : :
orfl41a.pep	CIDIGGGDLHTRIVWTQYGTLPHRVGDVQCRYRIVRLPQNADAPQGWQTVWQGARPRNKD : : : : : : :
orfl41-1	CIDIGGGDLHTRIVWTQYGTLPHRVGDVQCRYRIVRLPQNADAPQGWQTVWQGARPRNKD : : : : : : :
orfl41a.pep	SKFALIRKTGENI :
orfl41-1	SKFALIRKTGENI :

Homology with a predicted ORF from *N.gonorrhoeae*

ORF141 shows 95% identity over a 140aa overlap with a predicted ORF (ORF141ng) from *N.gonorrhoeae*:

60 orfl41.pep	DFGISPVYLWVAAAFAKHLSSPWAADSYDVA 30 : : : : : : :
orfl41ng	WNPAEPAVYTAVEALAGSPTPLV AHLFGQTDFGIPPVYLWVAAAFAKHLSSPWAADPYDA 126

-341-

	orf141.pep	RFAGGVFFAVIGLTS CGFAGFNFLGHRRGRXVVLILIGCIGLIPVVAHFLNPA AAAAFAAAGL	90
	orf141ng	RFAGGVFFAVIGLTS CGFAGFNFLGHRRGRXVVLILIGCIGLIPVVAHFLNPA AAAAFAAAGL	186
5	orf141.pep	VLGHGSILARRRVIASFL LGTGTWMLS LAAAYPAAFALMLPLPVLMFWRPWQSRRMLTA	140
	orf141ng	VLGHGSILARRRVIASFL LGTGTWMLS LAAAYPAAFALMLPLPVLMFWRPWQSRRMLTA	246

An ORF141ng nucleotide sequence <SEQ ID 599> was predicted to encode a protein having amino acid sequence <SEQ ID 600>:

10	1	MPSSEAVSARG CTEYLHLHAI RPFLLTLMIT YTFPDARPPA KTHEKFWLLL	
	51	LMAFAWLNG VFSHDLWNPA EPAYVTAEV LAGSPTPLVA HLFQGKTDGFI	
	101	PPVLYLVAAG FKHLHSPWA HPYDAAFAG VFVRAIGLTS CGFAGFNFLG	
	151	RHGRHSVVL HIGCIGLIPV AHFFNPNAAA FAAAGLVLHG YSLARRRVIA	
	201	ASFPLTGWTM LMSLAAYPA AFALMLPLPV LMFRFPWQSRLMLTAVASL	
15	251	AFAFLPMVTY PLLAKQTQPA LFQWLNVYH FGTFGVRH QRAFSLFHYL	
	301	KNLWFLWFPGP LPЛАWTVCR PLFSTDVGII LGIWMMLAVL VLAFLNPQRF	
	351	AMNYGWPAKL AERAAYFSFY YVDPIDPM AVAVLFPTLW LWAJTRKRNIR	
	401	GRQAVTNWAA GTVLWALMA KVTLFWLDA KSHAPVVRSM EASFSPELKR	
20	451	ELSJDGIECIG IGGGDGLBTRI WVTQCYGTLPH RVGDVRCRYR IVRLPQNADA	
	501	PQGWQTWQG APRRNKDSKF ALIRKIGENI LKTTD*	
	551		

Further work revealed the following gonococcal DNA sequence <SEQ ID 601>:

25	1	ATGCTGACCT ATACCCTGCC CGATGCGCG CGGCCCGCCA AAACCCACGA	
	51	AAAAACCGTC CTGCTGCTGT TGATGCGCTG TGCCTGCCTG TGGCCCGGG	
	101	TGTTTCCCCA CGATTGTTGG AATCTCGGCC AACCTGCGGT CTATACCGCC	
	151	CTGCGAACAC TGCGCAGCAG CCCACACCCC TTGGTTGCCCT ATCTGTCGG	
	201	TCAACAGCAT TTCCGCGCAT CGCCCGCTGTGA TCTTGGGGTT GCGCCCGCAT	
15	251	TCAACATTG CCTGTCGCCG TGGCGACGCC ACCCGTAGTA TGCGCGACGC	
	301	TTGCGAGGG TATTGTTGC CGTTATCGGA CTGACTTCTT CGCGCTTTGC	
	351	CGGTGTCAC ACCTTGGCGCA GACACCAAGG CGCACGGGTT TTAAATATCC	
	401	ATATCGCGCT ATATCGCGCT ATTCCCGGTT CCCATTTCCT CAATCCcgcc	
	451	ggccggccgc tGGCCGGCGC CGGACTGGG CTGCaaggct actcgctgGC	
	501	ACGCCGGCGC GTGATgtccg ccttcttccct GTCTCGTACG GGTGAGACGT	
30	551	TGATGTCGCT GCGCGCACGT TATTCGGCGC CGTTGGGGCT GATGCTGCC	
	601	CTGGCGCTGC TGATGTTTT CGTGTCCGGG CAAAGCAGGG GTTGTGATGT	
	651	GACCGCAGTC GCTCGCGCTG CCTTGTGCCCT CGCGCTTATG ACCGTTTAC	
	701	CGCTGCTtctt gCAAAAGCAG CAGCCGGCGC GTTTGGCGCA ATGGCTAAC	
	751	TATACGTTTCTT CGCTGAGCTT cggcgccgtt CGGCACatTC AGAggCatT	
40	801	Cagggtttt cactatctgt AAAatctgtc ttggttcgca ccggccgggc	
	851	TGGCGCTGGT GTTGTGGACG GTTGTGGCGCA CACGCGCTGTT TTGACCGAC	
	901	TGGGGGATTTC TGCGGATCTTG TTGGATGCTC GCGGTTTGG TGCTGCTCG	
	951	CTTAAATCTCA CGCGCTTTTC AAGACACACCT CGCTGGCGC CTGCGCGCC	
	1001	1001 TGCGCTGTG CGCGCGCGCG CAACTGACAA GCCTGAGGG CGGGGGGG	
45	1051	GCCTTGTGCA ACTGTTGCG CATTATGGCC TTCCGGCTGT TTGGCGCTGT	
	1101	1101 CCTGTGGACG GCGTTTTCTG CCTAGTAATTA CGCGCTGGCC CGCAACGCTT	
	1151	CCGCGACCGC CGCTCTACTC AGCGCCGTATT ACGTTCCCGA CATCGATTC	
	1201	1201 ATTCCGATGTC CGGTGTTGGCT ACTGTTACAA CCTCTTGCGC TGTTGGCGAT	
	1251	1251 CCTTAACTCA CGCGCTGGCT TTGGATGATA CGCTGTTCT CGCGTGGCT	
50	1301	1301 GCGTGTACCTT GACCTGGGGCT TTGGATGATA CGCTGTTCT CGCGTGGCT	
	1351	1351 GACCGCAGGC AAAGCACAGC CGCCGCTGGC CGGAGTATGG AGGACATCGTT	
	1401	1401 TTCCCGGGAA TTAACACGGC AGCTTTCAGA CGCGATCGAG TGATGCGCA	
	1451	1451 TAGCCGGCGC CGACCTGGAC ACGGGGATTG TTGGAGGGCA GTACGGCACA	
	1501	1501 TTGGCGGACG CGCTCGCGCA TGCCCTTGC CGCTTACCGTA TGTCCGGCT	
55	1551	1551 GCCCCAAAC CGGGATGGCG CGCAAGCGCT GCAGACCGTC TGGCAGGGT	
	1601	1601 OGCGCCCGCG CAACAAAGAC AGTAAGTTTG CACTGATACG GAAAATCGGG	
	1651	1651 GAAAATATAT TAAACACAC AGATGA	

This corresponds to the amino acid sequence <SEQ ID 602; ORF141ng-1>:

60	1	MLTYTPDAR PPAKTHKEFW LLLLMAFWL WPGVFSHDWL NPAEPAVYTA	
	51	VEALAGSPTP LV AHLFGQTD FGIPFVYLVW AAAFHKHLLSP WAADPYDAAR	
	101	FAGGVFFAVIGLTS CGFAGFNFLGHRRGRXVVLILIGCIGLIPVVAHFLNPA	
	151	AAAFAAAGGLV LHGSYLAGRRL ASIAASFLGT GTLMSLAAAG YPAAFALMLP	
	201	LPMLMFRPW QSRRMLTAV ASIAALPALPM TVPLPLIAKT QPALFAQWLN	
	251	VHVFGTFRGVW RHIQRAFSLF HYLKNNLWFA PGGLPLAWT VRTRLFSTD	
	301	WGILGIVNM LAVVLLAANP QRQCDNLVWL LPPLALFGAA QLDSLRRGAA	

351 AFVNWFQIMA FGLFAVFLWT GFFAMNYGWP AAKLAERAAYF SPYYVPDIDP
 401 IPMAAVLPLT PLWLWAIATRK NIGRQAVTN WAAGVTLTWA LLMTLFLPWL
 451 DAAKSHAPV RSMEASFSPE IRRRELSDGIE CIGIGGGDLH TRIWTCYGT
 501 LPHRGVDRC RYRIVRLPQN ADAPQGWQTV WGQARPRNKA SKFALIRKIG
 551 ENIILKTDX*

ORF141ng-1 and ORF141-1 show 97.5% identity in 553 aa overlap:

orf141ng-1.pep MLTYTTPDPARPKTHEKPWLILLMAFWLWPGVFSHDLWNPAEPAVYTAVEALAGSPTP
 orf141-1 MLTYTTPDPARPKTHEKPWLILLMAFWLWPGVFSHDLWNPAEPAVYTAVEALAGSPTP
 10 orf141ng-1.pep LV AHLFCQTDFGIPFPVYLWVAAAFKHLLSWAADPYDAARFAVGFFAVIGLTSCFGFN
 orf141-1 LV AHLFCQTDFGIPFPVYLWVAAAFKHLLSWAADSYDAARFAVGFFAVIGLTSCFGFN
 15 orf141ng-1.pep FLGRHHHGRSVVLHIGCIGLIPVAFHFLPAAAAAAGLVLHGYSLARRRVIAASFLLGT
 orf141-1 FLGRHHHGRSVVLHIGCIGLIPVAFHFLPAAAAAAGLVLHGYSLARRRVIAASFLLGT
 20 orf141ng-1.pep GTWITMSLAAAYPAAFALMLPLPVLMFTRPWQSRLMLTAVASLAFAPLMTVYPLLAKT
 orf141-1 GTWITMSLAAAYPAAFALMLPLPVLMFTRPWQSRLMLTAVASLAFAPLMTVYPLLAKT
 orf141ng-1.pep OPALFAQWLNHYHVGFTFGGVHRHJQRAFSLFLHYNKLWFLPAPPGLPLAVWTVCRTRLFSTD
 25 orf141-1 OPALFAQWLNHYHVGFTFGGVHRHJQTAFLSFPLYKLKNLLWFALPALPLAVWTVCRTRLFSTD
 orf141ng-1.pep WGILGIVVMMLAVLVLLAFLNPORFOONLWVLLPPALFGAAQOLDSLRGAAAFVNWFQIMA
 orf141-1 WGILGIVVMMLAVLVLLAFLNPORFOONLWVLLPPALFGAAQOLDSLRGAAAFVNWFQIMA
 30 orf141ng-1.pep FGLFAVFLWTGFTFAMNYGWPKALEAAYFSPYYVPDIDPIMPAVALFTPLWLWAIATRK
 orf141-1 FGLFAVFLWTGFTFAMNYGWPKALEAAYFSPYYVPDIDPIMPAVALFTPLWLWAIATRK
 35 orf141ng-1.pep NIRGRQAVTNWAAGVTLTWAALMFLPWLDAAKSHAPVVRSMEASFSPELKRELDGIE
 orf141-1 NIRGRQAVTNWAAGVTLTWAALMFLPWLDAAKSHAPVVRSMEASLSPELKRELDGIE
 40 orf141ng-1.pep CIGIGGGDLHTRIVWTQYGTLPIRVGDRCRYRIVRLPQNADAPGWOTWQGARPRNKA
 orf141-1 CIGIGGGDLHTRIVWTQYGTLPIRVGDRVQCRYRIVRLPQNADAPGWOTWQGARPRNKA
 45 orf141ng-1.pep SKFALIRKIGENILKTTDX
 orf141-1 SKFALIRKIGENIX

Based on the presence of several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 72

50 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 603>:

1 ..CAATCCGCCA AATGGTTATC GGGCAAACCT CTAGTCGGCA CAGCAATTGG
 51 GATACCGGG CAGATAAACG TTGGCCGCCA CCTGCATTAC GATATATTTA
 101 CGGCCGCCGC ATTGGAAAAG CCCGAATTTC TCCAATCAAG GAAATGGCA
 151 AGCGGTTTC AGGTAGGCTA TACGTTTTAA

55 This corresponds to the amino acid sequence <SEQ ID 604; ORF142>:

1 ..GSAKWLSGQT LVGTAIGIRG QIKLGGNLHY DIFTGRALKK PEFFQSRKWA
 51 SGFQVGYTF*

Further work revealed the complete nucleotide sequence <SEQ ID 605>:

1 ATCCATAATT CGGGTAGTCA GCGGACAGGA AAATACCAAG GAAATATCAC
 51 TTTCCTGCCA GACAATCCT TGCGACTGAG TGATATGTT TATGTAATT
 101 ATGCAGTTC GATTCGGGCT ACGCCGCGATG AGCAAAGTT TCACGGCCAT
 151 CGCAAAAGAC CGGGATCAAA CAATTAGCCC GTACATTATT CAGCCCCTT
 201 CGGTAAATGC ATACGGCAT TCACATCACAA TGCGTACCTT TACCATCAGG
 251 CAGTTCCCGC ATTATCGGAA GTCTATGACA ATAATGGAAA AAGTTACAT
 301 ACTGATTCG CTTCTCACCG CGTGTGTTAT CTCATGCCA AAGCCAAAC
 351 CTATCTCGGT GTAAAAGCTG GGATGAGGGG AACAAAAGCT TACATTGATG
 401 ATGCCGAAC TACTGTACAA CGGCCAAAAA CTGGGGTTG GTTCGCAGAA
 451 CTTCCTCCACA AAGAAATATG CGGTGCGAGT ACGCCGAGATT TAAAGTTGAA
 501 ATATAAACGC GGCACCGGCA TGAAGATGCG TCTGGGGCG CTCGAAGAG
 551 CCTTGGCGA AGGCAGCTCA CGTATGAAAATTTTGAGGGCG ATTCGGCTGAT
 601 GTAATAACTC CTTTCAACCA CGGTAAACAG CTATTTGCC ATACACATC
 651 CGTTATCGAC CAAATGAAACA AAACCCCGCT AACATCGCAA GACAAACCTGG
 701 CTATCCGGGC ACACCAACCC GTACGTGCTC TCGACGTGAA ATAGGTTTG
 751 TCTGGCGNGC GGGGATGTTA TTGGCGCACR GATTGAGGT GCGAATTTPAA
 801 ACCAGGGCAT CAGCTTITATC TTGGGGCTCA TGTTAGGACAT GTTCAGGAC
 851 AATCCGCCAA ATGGTTATCC GGCCAAACTC TAGTCGGCAC AGCAATTGGG
 901 ATACCGGGC AGATAAAAGCT TGCGGCACR CTGCAATTAGC ATATATTTCAC
 951 CGCGCGCGA TTGAAAGAGC CGGAATTTC CCAATCAAGG AATAGGCCAA
 1001 GCGGTGTTCA CGTAGGCTAT ACGTGTTTA

This corresponds to the amino acid sequence <SEQ ID 606; ORF142-1>:

1 MDNSGEATC KYQGNITFSAA DNPLGLSDIMF VNYYGRSIIIG TPDEESFDCH
 25 50 RKEGGSNYY VHAPSFGKWW TWAFNHNGYR YHQAVGSLSE VYDYGKNSY
 101 TDGFNRLLY RDAKRTYLY VLKLMRRETKS YIDDAELTVO RRKTAGWLAE
 151 LSHKEYIGRS TADFKLKYKR GTGMKDRLA PEEAFEGETS RMKIINTASAD
 201 VNTPFIQIKQ LFAYDFTSVHA QWNKPTLTQSQ DKLAIQGHHT VRGFIDGEMSL
 251 SAERGWYMRN DLSWQPKFPGH VSGQSAKWLIS QTLVGTAAIC
 301 IRGQIKLGN LHYDIFTQRA LKQLGADVHG VSGQSAKWLIS QTLVGTAAIC

30 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.gonorrhoeae*

ORF142 shows 88.1% identity over a 59aa overlap with a predicted ORF (ORF142ng) from *N.gonorrhoeae*:

35	orf142.pep	QSAKWLSGOTLVGTAIGIRGQIKLGGNLHY	30
	orf142ng	RGWYWRNDLWSQFKPGHQLYLGADVGHVGSQSAKWLSGQTLAGTAIGIRGQIKLGGNLHY	313
40	orf142.pep	DIFTGRALKKPFFQSKRHWASCFQVGYTF	59
	orf142ng	DIFTGRALKKPFFQTKWVTFQVGYSF	342

The complete length ORF142ng nucleotide sequence <SEQ ID 607> is:

1 ATGGATAATT CGGGTAGTCA GCGGACAGGA AAATACCAAG GAAATATCAC
 51 TTTCCTGCCA GACAATCCT TGCGACTGAG TGATATGTT TATGTAATT
 101 ATGCAGTTC GATTCGGGCT ACGCCGCGATG AGCAAAGTT TCACGGCCAT
 151 CCCAAAGAC CGGGATCAAA CAATTAGCCC GTACATTATT CAGCCCCTT
 201 CGGTAAATGC ATACGGCAT TCACATCACAA TGCGTACCTT TACCATCAGG
 251 CGTTTCCCG ATTATCGGAA GTCTATGACT ATAATGGAAA AAGTTACAC
 301 ACTGATTCG CTTCTCACCG CGTGTGTTAT CTCATGCCA AAGCCAAAC
 351 CTATCTCGGT GTAAAAGCTG GGATGAGGGG AACAAAAGCT TACATTGATG
 401 ATGCCGAAC TACTGTACAA CGGCCAAAAA CTGGGGTTG GTTCGCAGAA
 451 CTTCCTCCACA AAGGAATATG CGGTGCGACT ACGCCGAGATT TAAAGTTGAA
 501 ATATAAACAC GGCACCGCA TGAAGATGCG TCTGGGGCG CTCGAAGAG
 551 CCTTGGCGA AGGCAGCTCA CGTATGAAAATTTTGAGGGCG ATTCGGCTGAT
 601 GTAATAACTC CTTTCAACCA CGGTAAACAG CTATTTGCC ATACACATC
 651 CGGTATCGCA CAAATGAAACA AAACCCCGCT AACATCGCAA GACAAACCTGG
 701 CTATCCGGGC ACACCAACCC GTACGTGCTC TCGACGTGAA ATAGGTTTG
 751 CCTGGGGAGC GGGGATGTTA TTGGGGCAAC GATTGAGGT GCGAATTTPAA
 801 ACCAGGGCAT CAGCTTITATC TTGGGGCTGA TGTTAGGACAT GTTCAGGAC
 851 AATCCGCCAA ATGGTTATCC GGCCAAACTC TAGTCGGCAC AGCAATTGGG
 901 ATACCGGGC AGATAAAAGCT TGCGGCACR CTGCAATTAGC ATATATTTCAC
 951 CGGGCGTCCA TTGAAAGAGC CGGAATTATT TCAGACGAGG AATATGGTAA

1001 CGGGGTTTCA GGTGGGTTAT TCGTTTGAA

This encodes a protein having amino acid sequence <SEQ ID 608>:

```

1 MONGSEATW KYQONITFSA DNPFGLSDMF VNYVSGSLSE TPDDENFQD
5 RKEGGNNSYA PFAPEFGKW TWAFNHYGK YHQAVSGLSE VYDVNGKSYN
10 TDFGPNRLLY RDAKRKTYLS VKLNWTRETKS YDIAELTVO RRKTGTWLAE
15 LSHKGYIGRS TADFLKLYKH GTGMKDLARLA PEAGEGMTS KVIITASAD
20 VNTTPQIQKQ LEAYTDSVH QWNKNTPLTSQ DKLAIQGHHT VRGFGEMLS
25 PAERGWYWRN DLWSQSKWPGK QLYLGWVHGP QKLSAQSKWPGK QGLTAGTAIG
30 IRQIQLKGLMN LYHDYTFGR A LKHPKEYQTQ KWVQTFQWVGT SF*

```

- 10 The underlined sequence (aromatic-Xaa-aromatic amino acid motif) is usually found at the C-terminal end of outer membrane proteins.

ORF142ng and ORF142-1 show 95.6% identity over 342aa overlap:

```

15 orf142-1.pep MDSGSEATGKYQQNITFSDNPGLSDMFVYNYGRSISGGTDEEFSDFGRHKEGGNNYVA
orf142ng-1 MDSGSEATGKYQQNITFSDNPGLSDMFVYNYGRSISGGTDEEFSDFGRHKEGGNNYVA

orf142-1.pep VHYSAFPFGKWTWAFNHNNGYRHQAVSGLSEVVDYNGKSINTDFGNRLLYRDAKRTYLG
orf142ng-1 VHYSAFPFGKWTWAFNHNNGYRHQAVSGLSEVVDYNGKSINTDFGNRLLYRDAKRTYLG

20 orf142-1.pep VKLWMRRETTSKYIDDAELTQRQRTAGWLAEIHLKEYIGRSTADFKLKYKRGTMKDALAR
orf142ng-1 VKLWMRRETTSKYIDDAELTQRQRTAGWLAEIHLKEYIGRSTADFKLKYKRGTMKDALAR

25 orf142-1.pep PEEAFGETSRMKIITASADVNTPFQIGKQLFAYDTSVHAQWNKPTLTSQDKLAIIGGGHT
orf142ng-1 PEEAFGETSRMKIITASADVNTPFQIGKQLFAYDTSVHAQWNKPTLTSQDKLAIIGGGHT

30 orf142-1.pep VRGFDFGEMSLSAERGWYWRNDLWSQWPFGPHQLYLGYADVGHVGSQSAKWLSGQTGLWGTIAIG
orf142ng-1 VRGFDFGEMSLPAERGWYWRNDLWSQWPFGPHQLYLGYADVGHVGSQSAKWLSGQTGLWGTIAIG

35 orf142-1.pep IRQGK1KLGCNLHYID1FTGRALKPPEFFPSKRKA6GFOVGYTF
orf142ng-1 IROGK1KLGCNLHYID1FTGRALKPPEFFPSKRKA6GFOVGYTF

```

In addition, ORE142ng is homologous to the HscB protein of *E. chrysanthemi*.

```

40      gi|1772622 (L39897) RecB [Erwinia chrysanthemi] Length = 558
        Score = 119 bits (29%), Expect = 3e-26
        Identities = 88/346 (25%), Positives = 151/346 (43%), Gaps = 22/346 (6%)
Query: 2  DNSGSEATGKYQGNNNTFSANPFGLSMDFYVNYGRSISGGTDEENFDGHRKEGGSNYYAV 61
        DNSG +TT+ Q N + +DN EFLG D ++++ G S + + D + G
        Sbjct: 230 DNSQKQSTGEQQLNGLSALAWNVEGLADQWEIAGSHS---SFRATDHSASLQLAG---- 280
45      Query: 62 HYSAPFGKWTMAFNHNHGGRYHQARVSGLESEVDYNGKSYNTDGFGRNLRYRDAKRKTLYSV 121
        +S P+G W +N++ RY + + G S F +R++RD KT ++
        Sbjct: 281 -FSMSPGYWNLNLYNQSQRNFTFWSHGTSDTHFRFLSRSRVFRDGTMKTAIRG 339
50      Query: 122 KLMWTRTEKSYIADDALTVQRRKTTGWLAEELKHYIGRSTADFKLKHYGKMDKALRAP 181
        R +Y+ + L R + + +H + +F Y G +
        Sbjct: 340 TFSQRTGNNYLNGSLPSSSRKLSSVSLGVNHSQKLWGLGTLNFNTYGRVNRWLGESETD 399
55      Query: 182 EAEEAFGETSRKMIWTASADVNTPFTQIGKQLFRTDSVHQQWNKTPLTSLQKLAIGHHVT 241
        +++ E + WT SA P Y S+ Q++ L +L + G +G ++
        Sbjct: 400 DKSADEPRAEFNKTWSASYYHPV--TDSITYLGSLYQYSARALYQSGEQSLTQLTGHGESSI 456
60      Query: 242 RGFGEMLSPRAGWYRNLLDLSWQPK---GHOQLYLG---DVGHVSGSQSAKWLGSQPLLAG 296
        RGF E _ RG YWRN-L+M Q+ G+ ++ A D GH+ + + LG G
        Sbjct: 457 RGF-RQEYTSGNRGAYWRNLELNWQAWLPV LGNTFMAAVDDGHLYNHKQDNSTAASLWG 515
65      Query: 297 TAIGRQGIKLGNNLHYIDTGRALKKPEYFOTKKNVTFQGVQYS 342
        -G+C-A- L + C+A- B+C- V+C+G- U+C- G+C-

```

Sbjct: 516 GAVGMTVASRW---LSQQVTVGWPISYPAWLQPDTMVVGYRVGLSF 558

On the basis of this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

5 Example 73

The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 609>:

1 ATGGCGGAGCA AATGGTCAGC AGTGGAAAGC TGCTTACTTG GggGCCACCC
 51 GGCAGCATCG ATTACCGGCTT GAAACCTGTG TACCGCTTTGC AAAACCTAGC
 101 ATTCTCTCAT GGCCTGATGAAA AGCGCTCATG AGGACGCATC ATTTCWCGG
 151 ACAGCGAACCTT GGGCTTGTGCT ATGGAAACATT TGTCCGGACCA CGTAGAAGCC
 201 TTATGGCCATC ATTCGGACGGC TCCTGATCTT CCGAACGGCA ATTTCATCA
 251 TGAGGGCCGA GAAGAGTCTT GGTTGTTGGC GGCACAGACTG GCACAGATGG
 301 AAAAGAAATA CGGCCCTGTC ATTAGAACACA

This corresponds to the amino acid sequence <SEO ID 610: ORF143>:

1 MRTKWSAVRS CTTWADTADID TALNLLYRLQ KLEFLYGDEN GHSDGINLXD
51 EQLEMLMEQL SGSSGKALLVD RNLGLYLANAN FHHEAAEELG LLAEEVAQME
101 KKYRLLIKNN ..

Further work revealed the complete nucleotide sequence <SEQ ID 611>:

20	1	ATGGAACTCAA	CATTCTAC	ACAGACGAAAT	TATATGCCAA	GCGTACTCC
	51	TCCGGCTGCA	TTTTATCGCC	TATTCAGCG	GCCCCCCCAGT	GGCTGTAAAAA
	101	CTTGTGCGA	CAGCCTGTTG	AAAGCAGATG	CGGACGAATA	GTCAGCAGST
	151	GAGAACGAGG	TATCTTGCAG	GGACATCCGGC	GACATGCGATA	CGCTTGTGAG
	201	CCTGTGTTAC	CTTTGCGAA	AACCTGGAATT	CCTCTATGCG	GATGAAACRG
	251	GTCATCATCGA	CGGCATCAT	TGTGCGGAG	AGCAATTTGCG	TTGGTCGATG
	301	GACACANTCGA	CGGGCACCGG	TAAAGCGGTA	TTGGTCGATG	GGACAGCTTC
	351	GTATCTCTGC	AAACCCGAT	TCCATCATGA	GGCGCGGAGA	GAGTTGGGGG
	401	TTGGCGCCGG	AGAAATGCGCA	CAGATGAAA	AAAGATAATCC	GCTGCTGATI
	451	AAAGAACACC	TGTATATCGA	CATAATCGGT	TGGGGCGTTT	GGACGCTTCC
	501	CGGTCCAGACG	GAATTGACAT	TTTTCCCAT	TTATATCGCT	TCAACAAAT
30	551	TTATTTGGG	TAATCGCCGG	ATTCGGCGATT	TGGCCAGAAGA	GGCATTTGTT
	601	ACTTTGGTAA	GGATTTTATA	CCGGCTTAC	ACGAACCCGG	TGTTA

This corresponds to the amino acid sequence <SEQ ID 612; ORF143-1>:

1 MESTLSQLAN LYPRLTTPAGA FYAVSSDAPS AGKTLHSSL KADADEMVSS
51 EKLTTWATA DILDTALNRL RLIQKELFLPYG DENGHSGDIN SEZDQLPMLR
101 LPLGSGSKKA LUDVRNLGLYLA NANFHHEAB ELGLLAAVEA CMKEYYRLI
151 KNNLYLIRRNN WGVGCPDSGGS ELTFPPFLYIG STKFILVIGG IPDLGKRAFV
201 TULVRLYIRNN SNRV*
35

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

40 ORF143 shows 92.4% identity over a 105aa overlap with an ORF (ORF143a) from strain A of *N. meningitidis*:

45	orf143.pep	MRTKWSAVRSCWTADIDTALNLLYRLQKLEPL : GAFYAVSSDXPSAGKTLLHSLLKADADEMVSSEKSLLTWAXTADIDTALNLLYRLQKLEPL 20 30 40 50 60 70					
50	orf143.pep	40	50	60	70	80	90
		YGCENGHSDGTLNXLDEQLPLMEQDLSGSGKALLVNRNGLYLANANFHHEAELGLLAAE					

-346-

orf143a	YGDENGHSDGINLSDEQLPLLNEQL3GSGKALLVDRNGLYLANANFHHEAAEELGLAAE 80 90 100 110 120 130
5	100 110 orf143.pep VAQMEKKYRLLKNN
	VAQMEKKYRLLXKINNLYINNNNAWGVCDPSGQSELTFFPLYIGSTKFILVIIGGIPLDGKEA 140 150 160 170 180 190

The complete length ORF143a nucleotide sequence <SEO ID 613> is:

10	1	ATGGAAATCAA	CANTTTCACT	ACAAGCAAT	TTATATCNC	GCCTGACTCC
	51	TGCGGCCGTC	TTTTATGCCG	TATCCGCGGA	TGCCCCAGGT	GCGGTAAAA
	101	CTTGTGTC	CAGCGCTTGTG	AAAGCCGATG	CGGCAGAACAT	GTNNAGCATG
	151	GAGAACGTC	TTACCTGGCG	GGANAAACGCCG	CGCCCTTGGA	CGCCCTTGGA
	201	CTCTGTTGAC	CTTTGTCGAA	AACCTGGAATT	CCTCTATGGC	GATGAAAACCG
15	251	GTCATTCAGA	GGCCGATCAG	TGTTGGCGAGG	AGCAATGGC	GTGTCGATG
	301	GACATCTG	CGCCGCCGCG	TAAAGGCC	TTGGCTGRTG	GGACAGCTTC
	351	GTATCTTGC	AAACGCCAAT	TCCATCATGA	GGCGCGGAA	GAGTTGGGGT
	401	TGTTGGCGGC	AAAGATGCGA	CAGATGCGGAA	AAAGATAACCG	TCGCTNNNAT
	451	AAAGAACACC	TGTATATCAGA	CAATAGCGGT	TGGGGCGGTT	GGGATCTTC
20	501	CGGTGACACC	GAATTGCTT	TTTTCCCAAT	TTGATATCGT	CCAAACAAAT
	551	TTATTTTGTG	TATCCGGCGC	ATTCGGGATT	TGGGAAAGAG	GGCATTTGTT
	601	ACTTTGTTAA	GGATNTTATA	CCNNCGGTTA	CAGCCAGCGC	GTGTAACAACT
	651	TGGGAGGAG	GANGGGTTAT	CGCACGTTA	TTGA	

This encodes a protein having amino acid sequence <SEQ ID 614>:

25 1 MESTKSLQAN LYXRLTPAGA FYAVSSDXPS AGKTLHLSL KADADEMVSS
51 51 EKLTLWXTA DSDTLLRQ LQLQKQDNG DEHNSQD LQDLSQPLMM
101 101 ELGLSGKGAL LVDVRNLGLYLA NANFHHEAAE ELGLLAAEVA QMKEYRKLXI
151 151 KNNLYINNNA WGVCPDSQNS ELFTFLYLIG STKFILWIGG IPDLGKEAFV
201 201 TFLTRXLXYXL QPQRVLPLGRG XGLCNSY*

30 ORF143a and ORF143-1 show 97.1% identity in 207 aa overlap:

	orf143a.pep	MESTKSQIQLANLYXXLTPAGAFYAVSSDSDPAGKTLHSLKLADADEMVSSERSEKLTLWXTA
	orf143-1	MESTLSLQANLYPRLTTPAGAFYAVSSDAPAGKTLHSLKLADADEMVSSERSEKLTLWDTA
35	orf143a.pep	DIDTALNLNLLYRLQKLEFLYFLGDENGHSIGINLSDSEQPLLMEQ1LSGSGKALLVDRNGLYLA
	orf143-1	DIDTALNLNLLYRLQKLEFLYFLGDENGHSIGINLSDSEQPLLMEQ1LSGSGKALLVDRNGLYLA
40	orf143a.pep	NANFHHEAAEGLLAAEVAQMEKKYRXLKINNNLYINNNAWGVCDPSQGSELTFFFPLYIG
	orf143-1	NANFHHEAAEGLLAAEVAQMEKKYRLLKINNNLYINNNAWGVCDPSQGSELTFFFPLYIG
	orf143a.pep	STKFILWGGIPDLGEKAFTVLRVXLY
45	orf143-1	STKFILWGGIPDLGEKAFTVLRVILY

Homology with a predicted ORF from *N. gonorrhoeae*

ORF143 shows 95.5% identity over a 110aa overlap with a predicted ORF (ORF143ng) from *N.gonorrhoeae*:

An ORF143ng nucleotide sequence <SEQ ID 615> was predicted to encode a protein having amino acid sequence <SEQ ID 616>:

1 MRTKWSAVRS CSRADTADID TALNLLYRLQ KLEFLYQDEN GHSDGINLSD
 51 EQLPLLMEQL SGSSKGALLRND RNLGLYLANAN FHHESAEELG LLAEEVAQME
 101 KKYLLIRRN LYINNNNAWVG CDPGQSELT FFPLYIGSTK FLVIAIGPD
 151 LSGKGCICFG KDFIPPLQOOP RVKLGTGGIM QOLLISILED LNNFTSDIIA
 201 SAVISTDGLP MATMLPFSHNL SDRVGAISAT LLALGSRSVQ ELACGELEVQ
 251 MIRGKSGYIL LSQAGKDAVL VLVAKETGRL GLILLDDAKRA ARHIAEAI*

Further work revealed the following gonococcal DNA sequence <SEQ ID 617>:

1 ATGGAATCAA CACTTTCACT AACAAGGAAT TTATATCCCT GCCTGACTCC
 10 51 TGCGGTGCC TTTATTCGCG TATCCAGCGA TGCCCCCAGT GCCGGFAAAA
 101 151 CTTGTTGCGG CGAGCTTG AAAGCGGATG CGGAGCGAAGT GGTGACAGT
 151 GAGAACGTC TGCGCGCGG CACCGCGAC ATCGATACCG CTTTGACCT
 201 201 GTTGTAACCGT TTGCAAAAC TCGAAATTCT CTATGCGGAT GAAAACGGTC
 251 ATTCAAGACGG CATCAATTT CGCGACGAGC AATTGCCCCT CCTGATGGAA
 301 CAATTGTCCC CGACGGTAGA GGCAATTATTC GTCGATCGGA AGCGCTGTGA
 351 TCTGCGCACAC GCGCAATTCC ATCATGAGTC GCGCGGAAGAG TTGGGGTTGT
 401 TGCGCGAGC AGTCGACAC AGTGGAAAAGA AATACCGGCT GCTGATTAGG
 451 AACAACTGAT ATATCAACAA ATGCGTTG GCGCTTGGC ATCTTCCCG
 501 TCAGAGCGGAT TTGACATTCT TCCCATGTC TATCGGTCA ACCAAATTTA
 551 201 TTGGTTGTT CGCCGGCATT CGCGATTITGA GCAAAGAGGCC ATTGGTACT
 601 TTGGTAAGGA TTTTATACCG CGCTTACAGC AACCGCTGT AA

This corresponds to the amino acid sequence <SEQ ID 618; ORF143ng-1>:

1 MESTLSLQANLYPCLTPTAGA FYAVSSDAPS AGKTLLRSLL KADADEVVSS
 51 EKLLAADDAT IDTALNLYR LOKLEFLYGD ENGHSDGINL SDQLPLM
 101 OLSGSKALL VDRNGLYLAN ANFHHESAE LGLLAAEVQO MEKKYLLLR
 151 NNLYINNNAW GVCDPSQSE LTFFLYIIG TKFILVIAGI PDLSEKAFTV
 201 251 LVRILYRRYS NEV*

ORF143ng-1 and ORF143-1 show 95.8% identity in 214 aa overlap:

30	orf143ng-1.pep	MESTLSLQANLYPCLTPTAGA FYAVSSDAPSAGKTLLRSLL KADADEVVSS : : : : : : : : <td>orf143-1</td> <td>MESTLSLQANLYPRLTPTAGAFYAVSSDAPSAGKTLLRSLL KADADEVVSS : : : : : : : </td> <td>59</td>	orf143-1	MESTLSLQANLYPRLTPTAGAFYAVSSDAPSAGKTLLRSLL KADADEVVSS : : : : : : : 	59
35	orf143ng-1.pep	DIDTALNLLYRLQKLEFLYGDENGHSOGINLSDQEPLPMEQLELSGGKALLVDRNGLYLA : : : : : : <td>orf143-1</td> <td>DIDTALNLLYRLQKLEFLYGDENGHSOGINLSDQEPLPMEQLSGSGKALLVDRNGLYLA : : : : </td> <td>119</td>	orf143-1	DIDTALNLLYRLQKLEFLYGDENGHSOGINLSDQEPLPMEQLSGSGKALLVDRNGLYLA : : : : 	119
40	orf143ng-1.pep	NANFHHEAE LGLLAAEVQO MEKKYRLIIRNNWGVCDPSQSELTFFLYIG : : <td>orf143-1</td> <td>NANFHHEAE LGLLAAEVQO MEKKYRLIIRNNWGVCDPSQSELTFFLYIG : </td> <td>179</td>	orf143-1	NANFHHEAE LGLLAAEVQO MEKKYRLIIRNNWGVCDPSQSELTFFLYIG : 	179
45	orf143ng-1.pep	STKFILVIAGIPDLSKEAFVTLVRILYRRYSNRV 213 : <td>orf143-1</td> <td>STKFILVIAGIPDLSKEAFVTLVRILYRRYSNRV 214 : </td> <td>180</td>	orf143-1	STKFILVIAGIPDLSKEAFVTLVRILYRRYSNRV 214 : 	180

Based on the presence of the putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 74

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 619>:

1 ATGACCTTT TACAAACCTT GCAAGGTTTG CGAGACARATA AAATCTGTGC
 50 51 GTTTGCGATGG TTGCGTCTCC GCGCCCTTGA TGAAGAACCG TGACCCGAG
 101 101 CGGGCGCAAG CATGACCTTT AGCACCTGTC TTGGCACFCGT CCCCGTGGCTG
 151 ACCGTGAAGG TTGGCGCTGC TTGCGATTTTC CGCGTGTGTC ACCCGTGGTC
 201 201 GGATGCTGTC GTCTCTTCG TCAACCAAAAC CATTGTGCCG CA GGCGCG
 251 ACATGGTGTG CGACATPATC AATGCGTTCC CGCAGCGAGC GAACCGGCTG
 301 ACGGCAATCG CGACGCTGAT GCTGTTGCTT ACCTCGCTGA TGCTGATTCG
 351 351 GACGATAGAC AATACGTTCA ACCGCATCTG GCGGGTCAA wTCCAGCGT
 401 CGGTGGAT..

This corresponds to the amino acid sequence <SEQ ID 620; ORF144>:

```

1  MTFLQLRQLGL ADNKICAFAW FVVRRFDEER VPQXAASMTF TTLLALVLVPL
51 TVMVAVASIF PVFDRWSDF VSFVNQTIVP QGADMVFVYI NAFREQANRL
101 TAIGSVMVV TSMLMLIRTID NTFNRIWRVX XQRPWM...

```

5 Further work revealed the complete nucleotide sequence <SEQ ID 621>:

```

1  ATGACCTTT TACAACGTTT GCAGAGTTTG GCAGACAATA AAATCTGTGC
51 GTTGGCATGG TTCGCTCTCC GCGCCTTTGA TGAAAGAACGC GTACCGCAGG
101 CGGCGCAAGG CATGACGTTT ACAGGAGCTGG TGCGACTCGT CCCCGTGGCTG
151 ACCGCTATGG TGCGCGTFCG TTGATTTTCCCCTGTGCG ACCGCTGGTC
201 GGATCTGTGTC GTCCTCTCG CAACCAAAAC CATTGTCGCCG CAGGGCGGGG
251 ACATGGTGTG CGACTATPATC AATTCGCTTC GCGAGCACGG GAACCGCGTG
301 ACGGCAATGC CGACGGCTGAT GCTGGCTGTC ACCTCGTGTGA TGCTGATTG
351 GAGCATAGAC AATAGCTTC ACGCGATCTG GCGGGTCAAT TCCCCAGGTC
401 CGTGTGATGAT CGACTTCTC GCTCTATGGG CTTACTGTGAC GTTCTGGGGCG
451 CTGCTCTTGG GCGGGCGCAT TCCCTTATG TGCGCTCGG TAGCAGGATG
501 CGGCGCTGGC TAAGGGCGGC CGCACTGGTC GGGCGCGTTG CGAACCGGGG
551 CGRCGCTGAC CCTTCATGAC CTTGGCTGTG GGGGGCTGTA CGCTCTGTG
601 CCAACCGCTG CGCTGGCTGG CGCGCAGGGC TTGGCTGGGGG CTTGGCCAC
651 AGCGCTTGTG CTGGAACCGG CGGGCTCCCT CTTCACTGG TTATGGCA
701 ATTTGACGG CTACCGCTG ATTTACGGG CGTTTGCGC CGTGCCTGTT
751 TTTCTGTG TGCTGACCTT GTTGTGGACG CTGGCTCTGG CGCGCGCGGT
801 GCTGACTTCT TCACCTCTCT ACTGGCAAGG AGAACGCTTC CGCAGGGGCT
851 TCGACTCGG CGGAGCTTGTG GACGAGCTGTG TGAAAATCTT CGTCTCTCTG
901 GATGCGGGG AAAAAGAGG CAAAGCTTC CCTGTCAGG AGTTCAAGACG
951 GCATATCAAT ATGGGCTAACG ACAGAGTTGG CGAGCTTTTG GAAAAGCTGG
1001 CGGGCGAAGG CTACATCATC TCGCGCAGAC AGGGTTGGGT GTTAAAAGCG
1051 GGGCGCGATG CGATTGAGTT GAACGAACATC TTCAAGCTCT TGCTTFAACG
1101 TCGGTGGCT GTGGAAGAGG ATCATGTGAC CCAAGCTGTG GATGCGGTA
1151 TGACACGCTG TTTCGACAGT TGAACTATGA CGCTGGCAGA GTTGGACGCT
30 1201 CAGGGAAAA AACGGCGATA G

```

This corresponds to the amino acid sequence <SEQ ID 622; ORF144-1>:

```

1  MTFLQLRQLGL ADNKICAFAW FVVRRFDEER VPQQAASMTF TTLLALVLVPL
51 TVMVAVASIF PVFDRWSDF VSFVNQTIVP QGADMVFVYI NAFREQANRL
101 TAIGSVMVV TSMLMLIRTID NTFNRIWRVX SORFWMMQFL VYWALLTFGP
151 LSLVGQISEM VGSVQDAALI SAGAPQWSGAL RITAATLFTM LLLWGLYRFV
201 PNRVTPARQK FGVALATAFC LETARSLEPTW YMGNFGDYRS IYGAFAVAPF
251 FLLWNLWWL LVLLGAVLTL SLSYQGEAT RRGFDSSGRF DDVLLKLILL
301 DAAKEKGAL PKQEFRRHIN MGYDELGELL EKLARHGYYI SGRQGWVLKT
351 GADSIELNEL FKLEYFPRLP NERDHVNQAV DAVMTPCLQT LNMTLAEFDA
40 401 QAKKRQ*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N. meningitidis* (strain A)

ORF144 shows 96.3% identity over a 136aa overlap with an ORF (ORF144a) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
45	orf144.pep	MTFLQLRQLGADNKICAFAWFVVRRFDEERVPQXAASMTFTTLLALVLVPLTVNVAVASIF					
	orf144a	MTFLQLRQLGADNKICAFAWFVVRRFDEERVPQQAASMTFTTLLALVLVPLTVNVAVASIF	10	20	30	40	50
50	orf144.pep		70	80	90	100	110
	orf144a	PVFDRWSDSFVFSVNQTIVPVGADMVFVYINA 	120				
55	orf144.pep	PVFDRWSDSFVFSVNQTIVPVGADMVFVYINA 	70	80	90	100	110
	orf144a	PVFDRWSDSFVFSVNQTIVPVGADMVFVYINA 	120				
60	orf144.pep	NTFNRIWRVXXQRPWMN 					
	orf144a	NTFNRIWRVNSQRPWMMQFLVYWALLTFGFLSLGVGISEFXVGSVQDAALASGAPQWSGAL					

130	140	150	160	170	180
-----	-----	-----	-----	-----	-----

The complete length ORF144a nucleotide sequence <SEQ ID 623> is:

```

1 ATGACCTTT TACAACGTTT GCAAGGTTTG GCAGACATA AAACTGTGTC
5 51 GTTTGGATGG TTTCGTCGCC CGCGCTTITG TGAAAGAACGC GTACCCGAGG
10 101 CGGGCGCAG CATGACGTTT ACAGACATGT TGTCGACTGTG CCCCCGTGCTG
15 151 ACCGTGATTG TGGCGCTGCG TTGGATTTTTC CCCGTGTTGC ACCGNTGGTC
20 201 GGATCTGTC TTCTCTTCC TCAACCAAAC CATTGTGCGG CAGGGGGCGG
25 251 ACATGGTTNT CGACTATATC AATGCGTTC GCGGACAGGC GAACCGGGCTG
30 301 ACGGCATCG CGACGGTGTAT GCTGGTGTGTT ACCTCGNGA TGCTGATTG
35 351 GACGGATAGAC AATACGTTCA ACGGCATCTG CGGGGTCAT TCCACAGCTC
40 401 CGTGGATGAT CGACGTTCTC GTCTATTGGG CTTACTCTAC GTTTGGGGCG
45 451 CTGCTTGGG CGCTGGCAT TTCTCTTATN GTCGGCTCGG TACAGGATGC
50 501 CGCGCTGGC TCACTGGCG CGCAGTGGTC GGGCGCGTTG CGAACGGCG
55 551 CGACGGTGAN CTTCATGAGC CTTTGTGTGTT GGGGGCTGTA CGCTTCGTTG
60 601 CCAACACCTG TGTTCTCCGC CGCGCANGGG TTTCGCGGGG CTTTGGCAAC
65 651 AGCGCTCTGT CGTGAACACCG CGCGTTCGCC CTTTACTTGG TATATGGCA
70 701 ATTTCGAGGCT ACACGGCTCG ATTTCAGGNC CGTTTGCOCG CGTGGCGTTT
75 751 TTCTCTTGTG GCGTGAACCT GTTGTGACAC CTGGTCTGG CGGGCGCGGT
80 801 GCTGACTCTCTC TCACTCTCTC ACTTCGAGGG AGAACGCTTC CGCAGGGNCT
85 851 TCGACTCGGC CGGACGGTTT GACGACGTTG TGAACAACTCTC GCTGCTCTG
90 901 GATGCGGGCG AAAAGGAGG GNAAGCGCTTG CCTGTTCAAG AGTTCAGAGC
95 951 GCAATCATCAC ATGGCTCTACG ACAGAGTTGGG CGAGCTTTG GAAAAGCTGG
100 1001 CGGGCGAGCG CTACATCTAT TCCGGCAGAC AGGGGTTGGGT TTGAAAACAG
105 1051 GGGGGCGATT CGATGTTGTT GAAACGAACTC TTCAAGCTCT TCCTTACCG
110 1101 TCCGGTGGCT GTGGAAAGGG ATCATGTGAA CCAAGCTGTC GATCGCGTAA
115 1151 TTGATGGCTG TTTCGACACT TTGACATGAA CGCTGGCAGA GTTIGACGCT
120 1201 CAGGGAAAA AACACGAGCA ATCTGTA

```

This encodes a protein having amino acid sequence <SEQ ID 624>:

```

1 MTFLQLRQLGL ADNKICAFAW FVVRRFDEER VPQAAASMTF TLLALVPVL
30 51 TVVMVAVASIF PVDFRWSDDP VSFVNQTVIP QGADMVFDYI NAFRQANRL
101 TAIGSMLWV TSXMLIRITD NTFNRIWRVWV SQRPMMMQFL VYAWLTLTGP
151 LSLIGVISPK VGSVODALA SGAPQWSGAL RTAATLXFMFT LLLWGLYLYRXV
201 PNRFVPRKX FVGALATFC LETARSLEPTV YMGNFDGYRS IYGFPAVVPF
251 FILLNLWNLWT LVLGVALVTS SLYSWGGAEF RRXFDSRGRF DDVLKILLL
301 DAAQKEGXAL PVQEPRHRIN MGYDELGEEL EXKLARHGYYI SGRQGVWLKT
351 GADSBIENEL FKLFVYPRPL VERDHVNQAV DAVMMPCLQT LNMTLAEPFDA
401 QAKQQQS*

```

ORF144a and ORF144-1 show 97.8% identity in 406 aa overlap:

40	orf144a.pep	MTFLQLRQLGLADNKICAFAWFVVRRFDEERVPQAAASMTFTLLALVPVLTVMVAVASIF
	orf144-1	MTFLQLRQLGLADNKICAFAWFVVRRFDEERVPQAAASMTFTLLALVPVLTVMVAVASIF
45	orf144a.pep	PVFDWRSDSFVSFVNVQNTIVPQGADMVFDYI NAFRQANRLTAIGSVMVVTSXMLIRITD
	orf144-1	PVFDWRSDSFVSFVNVQNTIVPQGADMVFDYI NAFRQANRLTAIGSVMVVTSLMLIRITD
50	orf144a.pep	NTFNRIWRVNSQRPMMMQFLVYAWLTLFGLPSLGVGISFXVGSVQDAALASGAPQWSGAL
	orf144-1	NTFNRIWRVNSQRPMMMQFLVYAWLTLFGLPSLGVGISFXVGSVQDAALASGAPQWSGAL
55	orf144a.pep	RTAATLXFMFTLLLWGLYLYRXVPRXEVPRXAFVGALATAFCLETARSLEPTWYNGNFDGYRS
	orf144-1	RTAATLTFMTLLLWGLYLYREVPNRFPVPRQAFVGALATAPCLETARSLEPTWYNGNFDGYRS
60	orf144a.pep	IYGFPAVVPFFLWLNLWNLWTLVLGVALVTSLSYQGEAFRRXFDSRGRFDDVLKILLL
	orf144-1	IYGFPAVVPFFLWLNLWNLWTLVLGVALVTSLSYQGEAFRRXFDSRGRFDDVLKILLL
65	orf144a.pep	DAAQKEGXALPVQEPRHRINMGYDELGEEL EXKLARHGYYI SGRQGVWLKTGADSIENEL
	orf144-1	DAAQKEGXALPVQEPRHRINMGYDELGEEL EXKLARHGYYI SGRQGVWLKTGADSIENEL
	orf144a.pep	FKLFVYPRPLVERDHVNQAVDAVMMPCLQTLNMTLAEFDAQAKKKQ 408
	orf144-1	FKLFVYPRPLVERDHVNQAVDAVMTPCLQTLNMTLAEFDAQAKKRQ 406

Homology with a predicted ORF from *N.gonorrhoeae*

ORF144 shows 91.2% identity over a 136aa overlap with a predicted ORF (ORF144ng) from *N.gonorrhoeae*:

5	orf144.pep	MTFLQRLQGLADNKICAFAWFVVRFFDEEERPQXAASMTFTLLALVPVLTVMVAVASIF	60
	orf144ng	MTFLQCWQGSADNKICAFAWFVIRRFSEERVPQAAASMTFTTLLALVPVLTVMVAVASIF	60
10	orf144.pep	PVFDRWSDFSVFVNQTCIVPXGADMVFDYINAFAREQANRLTAIGSVMVVTSMLIRTID	120
	orf144ng	PVFDRWSDFSVFVNQTCIVPQGADMVFDYIDAFRDQANRLTAIGSVMVVTSMLIRTID	120
	orf144.pep	NTPNRRIWRWVXXQRPWM	136
15	orf144ng	NAFPNRRIWRWVNTQRPWMQFLVYWALLTFGPLSLGVGISFMVGSVQDSVLSGQAQQWADAL	180

The complete length ORF144ng nucleotide sequence <SEQ ID 625> is predicted to encode a protein having amino acid sequence <SEQ ID 626>:

1	MTFLQLCWQGS ADNKICAFAW FVIRRFSSEER VPQAAASMTF TTLLALVPVL
51	TVMVAVASIF PVFDRWSDF SVFVNQTCIV PXGADMVFDYI DAFRDQANRL
101	TAIGSVMILVV TSMLIRTID NAFNRRIWRVN TORPWWMQLF VVWALLTFGP
151	LSQLQGISEM VGSVQDSVLS SGAQWADAL KTAARLAFLMT LLLWGLYRFV
201	PNRFVPARQA FVGALITAFC LETARFLFTW YMGNFDGYRS IYGAFAAVPF
251	FLLWNLNLLT LVLGAVLTS SLSYWGGEAF RRGFDSRGRF DDVLKILLL
301	DAAQKEGRRTL SVQEFRRHIN MGYDELGEEL EKLARYGYIY SGRQGVNLKT
351	GACDIELSEL FKLFVYRPLP VERDHVNQAV DAVMTPCLOT NNMTLAEFDA
401	OAKKKQQQS

Further work revealed the following gonococcal DNA sequence <SEQ ID 627>:

1	ATGACCTTT TACAACCGTG GCAAGGTTTG GCGGACACATA AAATCTGTGC
51	ATTGGCATG TGTCGATCC CGCGTTTCAG TGAAGAGCCG GTACCCGCCAG
101	CAGGGCGAGA CATGACGCTT ACGACACTG TGCGCACATCG CCCCCGTACTG
151	ACCGTAACTG TGCGGCTGCG TGTCGATTTTG CCCGTGTTTG ACCCGCTGGTC
201	GGATGCGTTG GTCTCTCTTG TCAACCCAAA CATTGTCGGC CAGGGCGCGG
251	ATATGCTGTG CGCATRACTAC GACGGATTC GCGGATCAGCG AAACCGCTG
301	ACCCGATCGC GACGGCTGAT CCTGGTGTAT ACCTGCGTGA JCTGTGATTG
351	GACGATAGAC ATAGCGGTTCA ACCGCGATCTG GCGGGTTAAC AGCGAACGCC
401	CCTGCGATGAT CGACGTCTTC GTTATTGTTGG CGTTCGTCGAC TTTGGGGCT
451	TGTCGCTTGTG GTGTCGGCAT TTCTTTATG GTGCGGCTGG TTCAAGACTC
501	CGTACGCTCTC TCGCGGAGCG ACAAATGGGG GGACGCGCTTG AAGACGCGCG
551	CAAGGCTGGG TTTCATGCGC CTTTCTGCTG GGGGGCTGTA CGCTTCTGTG
601	CCCAACCGCT CGTGTGCGCC CGCGCAGGCTT TTGGTCGGAG CTTTGGATTAC
651	GGCATTCCTG CTAGGAGACG CGAATTTCTC GTTCACTCTG TATATGGCA
701	ATTCGCGCTG CTACCGCTCG ATTACGGCTC CATTGGCGCC CGTGCCTGTT
751	TTCCTCTGTT GGTAAACACT GCTGTGGAAC CTGGTCTCTG CGCGGGCGGT
801	GCTGACTTCTG CGTCGCTCTT ATTGGCAGGG CGAGGGCTTC CGCAGGGGAT
851	TGCACTGCGC CGGAGCGTTT GACGACGCTG TGAAGAACCTCT GCTGCTTCTG
901	GATGCGCGGAG CGGAACCCCTG TCCGTCAGG AGTTCAAGACG
951	GCATATCAT ATGGTTTACG ATGAATTTGG CGAGCTTTTG GAAAAGCTGG
1001	CGCGGTACCG CTATATCTAT TCCGGCAGAC AGGGCTGGGT TTGAAAACCG
1051	GGGGCGGATT CGATTTGAGT GAGCGAACCTC TTCAAGCTCT CGTGTGACCG
1101	CCCGTTGct gtggAAAGG ATCATGTGAA CCAAGCTGtc gATGCGTAA
1151	TGAcgcgtg TTGCAAGACT TTGACATGTA CGCTGGCGGA GTTTGACGCT
1201	CAAGcgcAAA AACACAGCA GTCTTG

This encodes a variant of ORF144ng, having the amino acid sequence <SEQ ID 628; ORF144ng-I>:

1	MTFLQRWQGL ADNKICAFAW FVIRRFSSEER VPQAAASMTF TTLLALVPVL
51	TVMVAVASIF PVFDRWSDF SVFVNQTCIV PXGADMVFDYI DAFRDQANRL
101	TAIGSVMILVV TSMLIRTID NAFNRRIWRVN TORPWWMQLF VVWALLTFGP
151	LSQLQGISEM VGSVQDSVLS SGAQWADAL KTAARLAFLMT LLLWGLYRFV
201	PNRFVPARQA FVGALITAFC LETARFLFTW YMGNFDGYRS IYGAFAAVPF
251	FLLWNLNLLT LVLGAVLTS SLSYWGGEAF RRGFDSRGRF DDVLKILLL
301	DAAQKEGRRTL SVQEFRRHIN MGYDELGEEL EKLARYGYIY SGRQGVNLKT

351 GADSIELSEL FKLGVYRPLP VERDHVNQAV DAVMTPCLQT LNMTLAEFDA
401 QAKKQQQS*

ORF144ng-1 and ORF144-1 show 94.1% identity in 406 aa overlap:

5 orf144ng-1.pep MTFLQRWQLGADNKICAFAMFWVIRRFSEERVPQAAASMTTLLALVLPLVLTVMWAVASIF
orf144-1 MTFLQRLQLGADNKICAFAMFWVIRRFDEERVPQAAASMTTLLALVLPLVLTVMWAVASIF

10 orf144ng-1.pep PVFDRWSDFSFSVFNQNTIVPGADMVFYDIAFDRQANRLTAIGSVMVLVTSIMLIRTID
orf144-1 PVFDRWSDFSFSVFNQNTIVPGADMVFYDIAFREQANRLTAIGSVMVLVTSIMLIRTID

15 orf144ng-1.pep NAFNRNIRWVNNTQRPFWMQMFLVYWALLTFGPLSLGVGISMWGSVQDSVLSSGAQNQWADAL
orf144-1 NTFNRIIRWVNNSRPFWMQMFLVYWALLTFGPLSLGVGISMWGSVQDAALASGAPQNSGAL

20 orf144ng-1.pep KTAARLAFTMLLGLYRFPVNRPVPARQAFVGALATFCLETARFLTWYMGNFDCGYS
orf144-1 RTAATLFMFTMLLGLYRFPVNRPVPARQAFVGALATFCLETARSLFLTWYMGNFDCGYS

25 orf144ng-1.pep IYGFAAVVPFFLWLNLWLTIVLGVAVLTTSSLSYWQEAFRRGFDSDRGFRFDVLLKILLLL
orf144-1 IYGFAAVVPFFLWLNLWLTIVLGVAVLTTSSLSYWQEAFRRGFDSDRGFRFDVLLKILLLL

30 orf144ng-1.pep DAAQKEGRYLVSQEFRRRHINNGYDELGELLEKLARHGYIYISGRQGWVLKTGADSIELSEL
orf144-1 DAAQKEKGALPVQEFEERRRHINNGYDELGELLEKLARHGYIYISGRQGWVLKTGADSIELSEL

35 orf144ng-1.pep FKLFVYVPLPVERDHVNQADAVMTPCLQTLINTLAEFDDAQAKKKQQQS
orf144-1 FKLFVYVPLPVERDHVNQADAVMTPCLQTLINTLAEFDDAQAKKKRQQ

On this basis of this analysis, including the identification of several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

35 Example 75

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 629>:

1 ..AGACACGCC GCGCCTCGG CATCGACACC GGCATCAACC CCGAACTGGAA
 51 AGCCCTGGG GAAACACTCC ATTACCAATTG GCAGGGCTTC CTCTGGCTCA
 101 GCACCATGAT GCTTCAGGAA ATTTCGCCCC TTCTGTACCT CGTCGCAAC
 151 ACCCGCGGA AATGGCTGG TGGCCCAAAGA CGCCAACACC TGCGCCAAG
 201 CCTGGTGGAA ACAGCGGAAAC AGCGCTGA

This corresponds to the amino acid sequence <SEQ ID 630; ORF146>;

1 .RHARRIRIDT AINPEALEALA EHLHYQWQGF LWLSTDMRQE ISALVILLQR
51 TRRKWLDAHE RQHLRQSLLE TREHG*

45 Further work revealed the complete nucleotide sequence <SEQ ID 631>:

1	ATGAAACCCCTT	CGCCAAAGCT	CGCCCTGTGC	AGCCGGCTTC	TCAACTCTTA
51	CGAACCTAC	CGCTAACGCC	GCTCTATCCA	CGCCGGCGC	CTCGCGCGGG
101	CGCTGTGTT	CGCCACCGCC	TCCGGCGCGC	TGCTCCACCT	CAAAACAGCC
151	GAGTGATAG	GGATGACGCT	CTTCGGCTC	CTCTGGCATG	TCCAGTCTTCA
201	AGGGGGGATT	TACCTAACGA	CGGTGGAAAC	TATGCTCGGC	AGGGATCAGG
251	GGCTGGGGCG	GGGTATTGGCC	GTTTATAGGC	TAACGACCGA	TTATTTCAC
301	GGCAACCTAC	TCTCTTACCT	CACCGCTGGC	AGCCGGACGC	TCTGGCGGCC
351	CTGGGGGGCC	GTCTGGCAAAT	ACCGCTTACGT	CTCTTATGGCT	CGACGGCTGA
401	CGATGGTGTAT	GCTCATGGCG	ACACAGCGGA	GGAAATGGCT	CGACGGCGGC
451	CTTATGGCT	CCATGACGCT	CTCTCATGGC	GGGGCATCTT	CGACGGCGCC

5	501	CGCCAAACTG	CTGGCGGTGA	AATCCACACT	GATGTGGCGT	TTCATGCTTG
	551	CGGACACACT	GGCGCATG	AGCAAATGA	TTGCCGAAAT	CAGCAACGGC
	601	AGGGCGATGA	CCCGGGAACG	CCTCGAGGAG	AAACATGGCGA	AAATGGGCCA
	651	AATCAACGCA	CCGATGTC	AAAGCCGAG	CCATCTGGCC	GCACACATCGG
5	701	GGCAAGCGG	CATCAGCCCC	GCATGATG	AAGCCATGCA	GCAGGCCAC
	751	CGTAAATCG	TCACACACC	GGAGCTGTC	CTGACCCACG	CGGCAAGGT
	801	GCAATCTCCC	AAACTCAACG	GCAGCAAAT	CCGGCTGCTT	GACGGCACT
	851	TCACACTGCT	CCAAACCGC	CTGACAAACAA	CGCTGCCCT	TATCAACGGC
10	901	AGACAGGCC	GGCGATCG	CATGACAC	GCACATCAAC	CGGAACTGGA
	951	AGGCTTCGCC	GAACACCTCC	ACTACCAATG	GCAGGGCTTC	CTCTGGCTCA
	1001	GCACCAATAT	GGTCAGGAA	ATTTCGGCC	TGTCATCCT	GCTGCAACGC
	1051	ACCCGGCCCA	AAATGCTGA	TGCCCCAGGA	CGCCAACAC	TGGCCCAAAG
	1101	CCTGCTTGA	ACACGGGAAC	ACGGCTGA		

This corresponds to the amino acid sequence <SEQ ID 632; ORF146-1>:

15	1	MNTSQNRNLV	SRWLNSRYSY	RYRRLIHAVR	LGGAVLFTA	SARLLHLQHC
	51	EWIGMTVFWV	LGMILQFGQAI	YSKAVERMLIG	TWIGLGAGLG	VLWLNQHYPH
	101	GNLIFYLTVG	TASALAGWA	VGNKNGVYPM	AGLITMCLIG	DNGSEWLDS
	151	IMRMNVVLIG	AAIAIAAAKL	LPLKSTLMWR	FMLADNLADC	SKMIAEISNC
20	201	RRMRTRERL	NNAKKMRQINA	RMVKSRSHIA	ATSGESRISP	AMMEAMQHAA
	251	RKIVNTELL	LTTAAKQLQP	KLINGSEIRLL	DRHFTLLQTD	LQQFVALINC
	301	RHARURIRDT	AINPELEALA	EHLHYQWQF	LNLSTNMQRQE	ISALVILLQR
	351	TRRKWLDAHE	RQHLRQSLL	REHGS*		

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

25 ORF146 shows 98.6% identity over a 74aa overlap with an ORF (ORF146a) from strain A of *N. meningitidis*:

	orf146.pep	RHRARRIRIDTAINEPELEALAEHLHYQWQF				
30	orf146a	KLNGSEIRLLDRHF	TLLQDQQTVALINGR	HARRIRIDTAINEPELEALAEHLHYQWQF		
		280	290	300	310	320
		40	50	60	70	
35	orf146.pep	LWLSTDIMRQEISALVILLQ	LRTRRKWLDAHERQHLRQSLL	ETREHGX		
		340	350	360	370	
	orf146a	LNWSTNMQRQEISALVILL	QTRTRRKWLDAHERQHLRQSLL	ETREHXS		

The complete length ORF146a nucleotide sequence <SEQ ID 633> is:

40	1	ATGAAACACT	CGAACCGCAA	CGGCCCTCGTC	AGCCGCTGGC	TCAACTCCTA
	51	CGAACCTAC	CGCTTACCGCC	GCCTCATC	CCCGCTCCG	CTCGGGGG
	101	CCGCTCTG	CCGCCACGCC	TCCGGCCCG	TGCTCCACCT	CACACACGGC
	151	GAGTGGATAG	GGATGACCG	TCTCGTGTG	CTCGGCATGC	TCAGATTTC
	201	GGGCTGGGAT	TACTCCAAAG	GGTTTGGAAAGC	TATGTCGCGC	ACGGTCATCG
	251	GGCTGGGGCG	GGGTTGGCG	GTTCGGTGGG	TGAAACAGCA	TTATTTCCAC
	301	GGCAACCTCC	TCTTCACCT	CCGGCTGGG	ACGGCAAGCG	CACGGCCGG
	351	CTGGGGGGCG	GTGCGCAAA	ACCGCTTACGT	CCCTATGC	CGGGGGCTGA
	401	CGATGTCGAT	GTCATCGGC	GACAAGGGCA	GCGAATGGTT	CGCACAGGGC
	451	CTGATGTCGG	CGATGACCAT	CCTCATCGC	CGGCCATCG	CCATCGCCG
	501	CGCCAAACTG	CTGGCGGTGA	ATTCACACAT	GATGTGGCGT	TTCATGCTTG
	551	CCGACACACT	GACCGACTCG	AGCAAATGA	TTGCCGAAAT	CAGCAACGGC
	601	AGGGCGATGA	CCCGCGAACG	CCTCGAAGAG	AAACATGGCGA	AAATGGGCCA
	651	AATCAACGCA	CGCATGTC	AAAGCCGAG	CCACCTGGCC	GCACACATCGG
	701	GGCAAGCGG	CATCAGCCCC	GCATGATG	AAGCCATGCA	GCAGGCCAC
55	751	CGTAAATTTG	CAACACCCAC	CGAGCTGTC	CTGACCCACG	CGGCGCAAGCT
	801	GCAATCTCCC	AAACTCAACG	CGAGCAAAT	CCGGCTGCTT	GACGGCACT
	851	TCACACTGCT	CCAAACCGC	CTGACACAA	CGCTGCCCT	TATCAACGGC
	901	AGACAGGCC	GGCGCATCG	CATGACAC	GCACATCAAC	CGGAACTGGA
	951	AGGCTTCGCC	GAACACCTCC	ACTACCAATG	GCAGGGCTTC	CTCTGGCTCA
60	1001	GCACCAATAT	GGTCAGGAA	ATTTCGGCC	TGTCATCCT	GCTGCAACGC
	1051	ACCCGGCCCA	AAATGCTGA	TGCCCCAGGA	CGCCAACAC	TGGCCCAAAG
	1101	CCTGCTTGA	ACACGGGAAC	ACAGTGTGA		

This encodes a protein having amino acid sequence <SEQ ID 634>:

```

      1  MNTSQNRNLV SRWLNSYERY RYRRLIHAVR LGGAVLFLATA SARLLHLQHG
      5  EWIGMTVFV LGMLOFGQAI YSKAVERMLG TVIGLGRAGLG VLWLNQHYEH
     101  GNLFLVLTVG SALAGWAIA VGKNGVPMRL RGLTMCLIG DNGSEWFDSG
     151  LMRAMVLIG AAIAIAAAKL LPLKSTLWR FNLADNLTD C SKMIAEISNG
     201  RRMTTERLEE NNAMKRQINA RMVKRSRSHLA ATSGESRISP AMMEAMQHAA
     251  RKIVNTELL LTAAKQLSP KLNQSEIRRL DRHFTLLQTD LQQTVALING
     301  RHARRIRIDT AINPELEALA EHLHYQWQGF LWLSTNNRQE ISALVILLQR
     351  TRRKWLDAHE QHQHLRQSLE TREHS*

```

10 ORF146a and ORF146-1 show 99.5% identity in 374 aa overlap:

orf146a.pep	MNTSQNRNLVSRWLNSYERYRYRRLIHAVRLGGAVLFATASARLLHLQHGEWIGMTVFV
orf146-1	MNTSQNRNLVSRWLNSYERYRYRRLIHAVRLGGAVLFATASARLLHLQHGEWIGMTVFV
orf146a.pep	LGMLOFGQAIYSKAVERMLGTVIGLGRAGLGVLWLNQHYFHGNLLFYLTVTASALAGWA
orf146-1	LGMLOFGQAIYSKAVERMLGTVIGLGRAGLGVLWLNQHYFHGNLLFYLTVTASALAGWA
orf146a.pep	VGKNGVPMRLAGLITMCLIGDNGSEWFDSGLMRAMNVILIGAAIAIAAKLPLKSTLWR
orf146-1	VGKNGVPMRLAGLITMCLIGDNGSEWFDSGLMRAMNVILIGAAIAIAAKLPLKSTLWR
orf146a.pep	FMLADNLTDCKSMIAEISNGRRMTRERLEENMAKMRQINARMVKRSRSHLAATSGESRISP
orf146-1	FMLADNLADCSKSMIAEISNGRRMTRERLEENMAKMRQINARMVKRSRSHLAATSGESRISP
orf146a.pep	AMMEAMQHAAHRKVIVTTELLLTAAKQLSPKLNQSEIRLLDRHFTLLQTDLQQTVALING
orf146-1	AMMEAMQHAAHRKVIVTTELLLTAAKQLSPKLNQSEIRLLDRHFTLLQTDLQQTVALING
orf146a.pep	RHARRIRIDTAINPELEALAEHLHYQWQGF LWLSTNNRQE ISALVILLQRTRRKWLDAHE
orf146-1	RHARRIRIDTAINPELEALAEHLHYQWQGF LWLSTNNRQE ISALVILLQRTRRKWLDAHE
orf146a.pep	RHQHLRQSLE TREHSX
orf146-1	RHQHLRQSLE TREHG

Homology with a predicted ORF from *N.gonorrhoeae*

40 ORF146 shows 97.3% identity over a 75aa overlap with a predicted ORF (ORF146ng) from
N.gonorrhoeae:

orf146.pep	RHARRIRIDTAINPELEALAEHLHYQWQGF	30
orf146ng	KLNQSEIRLLDRHFTLLQTDLQTAALINGRHARRIRIDTAINPELEALAEHLHYQWQGF	364
orf146.pep	LWLSTNDMRQEISALVILLQPLQRTRRKWLDAHERQHLRQSLE TREHG	75
orf146ng	LWLSTNDMRQEISALVIPLQRTRRKWLDAHERQHLRQSLE TREHG	409

An ORF146ng nucleotide sequence <SEQ ID 635> was predicted to encode a protein having amino

50 acid sequence <SEQ ID 636>:

```

      1  MSGVRFPSP PIPSTDPPSG SLCFFTPFLQ TASDMNNSQR KRLSGRWLNS
      5  YERYRRRLI HAVRLGGTVL FATALARLILQ LHGEWIGMT VEVVLMQI.QF
     101  CGAIYISNAVE RMLGTVGVLG AGLGVLWLNQ HYFHGNLLFY LTIGTASALA
     151  GWAAGKNGY VMLLAGITMC MLIGDNGSEW LDSSGLMRAMN VLIGAAIAIA
     201  BAKLLELKST IMWREMLADCSKSMIAE ISNGRRMTR ELEQNVMVKMR
     251  QINAMRVKSR SHLAATSGES RISPSMMEM QHAAHRKIVNT TELLITTAAK
     301  LQSPKLNQSE IRLLDRHFTL QTQDLQTAAN LINGRHARRI RIDTAINFEL
     351  EALAEHLHYQ WQGFLWLSTN MRQEISALVI PLQRTRRKWL DAHERQHLRQ
     401  SLELETREHG*

```

60 Further work revealed the following gonococcal DNA sequence <SEQ ID 637>:

1 ATGAACTCCT CGCAACCGAA AGCCCTTTC GgcGCTGGC TCAACTCCATA
 51 CGAACGCTac cGCCaccGCC GCCTCATACAA TGCCGTGCGG CTGGCGggaa
 101 ccgtCTGTG CGCACCGCGA CTGGCCCGc ACTCACCT CCAcacccgc
 151 gGAATGGATAG GGATgacCGT CCTCGTCGTC CTGGCATGc TCCAGTTCCA
 201 AGGCgcgat tActccaae cggtgAaAgc tTgctcggt acggatcatcg
 251 ggctgGGG GGGTTTGccg gTTTTATGGG TGAACACGCA TTAtttccac
 301 ggcaacCTcc ttcttccatc gaccatccgc acggcaagcg cactggccg
 351 ctGGCCGGCC GTCGCCAAAGA cggcgtaacg cctatcgct GCGGGGctgA
 401 CGATGTGCA tgcataccgc gACAACGCCA GCGAATGGCT CGACAGCGGC
 451 CTGATGCGCA CGATGACGCT CTCATCGGC CGCCGCATCG CCATTGCGGC
 501 CGCCAAACTG CTGGCGCTGA ATACCAACT GATGTTGGCT TTCTATGCTTG
 551 CGCACAACTC GCGCGACTGC AAGCAAATYGA TTGGCGAAAT CAGCAGAACGGC
 601 AGGCATGAGA CGGCCGAACG TTGGAGCACG AATATGGTC AAATGGGCCA
 651 AATCACCGCC CGCATGTCA AAAGCCGCA CGAACATCGG GCGACATCGG
 701 CGGAANGGCG CATCAGCCCC TCCATGATGG AAAGCCATGCA GCGACGCCAC
 751 CGCAAAATCG TCAACACCC CATGGCTGC CTGACACCCG CGCGCAAGCT
 801 GCAATCTCCC AAATCTCAGG CGACGGAATAT CGCGCTGCTC GACCGCACT
 851 TCACACTGCT CAAACCGCG CTGCAACAAA CGCGCGCTC CATCACCGC
 901 AGACACGCC CGCCGAICCG CATCGACACC GCGATCACCC CGAACATGGG
 951 AGGCCTCGCC GACACCTCC ACTACCAATG GCAAGGCTTC CTCTGGCTCA
 1001 GCACCAATAT GUGTCAGGAA ATTTCGGCC TGGTCACTCT GTGCAAAAGC
 1051 ACCGGCCCA ATAGCTGGAA TGCCCCGAA CCCAACACCC TGGGCCAAAG
 1101 CCTGCTTGA ACACGGGAAc ACGGCTGA

This corresponds to the amino acid sequence <SEQ ID 638; ORF146ng-1>:

25 1 MNSSQRKRLS GRWLNSRVRHRRHLIHAVR LGGTVLVFATA LARLLHLQHG
 51 EWIGWTFEVV LGMLOFGGAI YSNAVERMLG TVIGLGAGLG VLRNLNHYLFH
 101 GNLYLFTIG TASALAGWAA VGKNGYVPMI AGLTMCMLIG DNGSEWLDSG
 151 LMRAMWLVIG AAAIAIAAKL LPLKSTLMNR FPLADNLADC SKMIAEISNG
 201 RRMTRERLEQ NVVKMRGIRNA RMVKSRSHLA ATSGESRISP SMMEARMQRAH
 251 RKIVNTTEL LTAALKLQSP KLNGESEIRLL DRHFTLLQTD IQQTAALING
 301 RHARRIRIDT AINPELEALA EHLHQWQGF LWLSTMNRQE ISALVILLQR
 351 TRRKWLDAHE RQHLRQSLSL TREHG+

ORF146ng-1 and ORF146-1 show 96.5% identity in 375 aa overlap

35 orf146-1.pep MNTSQRNRLVSWRNLNSYFFRYRRLIHAVRLGGAVLFATASARLLHLQHGEWIGMTVFVV
 orf146ng-1 MNSSQRKRLSGRWLNSRERYRHRRLIHAVRLGGTVLFATALARLLHLQHGEWIGMTVFVV
 orf146-1.pep LGMLQFQGAIYSKAVERMLGTVIGLGAAGLGVWLNLQHYFHGNLLFYLTGTASALAGWAA
 40 orf146ng-1 LGMLQFQGAIYSNAVERMLGTVIGLGAAGLGVWLNLQHYFHGNLLFYLTGTASALAGWAA
 orf146-1.pep VGKNGYVPMIAGLTMCMLIGDNESEWLDSLGLMRAMNVLIGAAIAIAAAKKLPLKSTLMNR
 orf146ng-1 VGKNGYVPMIAGLTMCMLIGDNESEWLDSLGLMRAMNVLIGAAIAIAAAKKLPLKSTLMNR
 45 orf146-1.pep FMLADNLADCCKMIAEISNGRRMTRERLEENMAKMRQINARMVKSRSHLAATSGESRISP
 orf146ng-1 FMLADNLADCCKMIAEISNGRRMTRERLEENMAKMRQINARMVKSRSHLAATSGESRISP
 50 orf146-1.pep AMMEAMQHARHKIVNTTELLTTAAKLQSPKLNNGSEIRLLDRHFTLLQTDLQTVALING
 orf146ng-1 SMMEARQHARHKIVNTTELLTTAAKLQSPKLNNGSEIRLLDRHFTLLQTDLQTVALING
 55 orf146-1.pep RHARRIRIDTAINPELEALAELHLYQWQGFWLSTNNRQEISALVILLQTRRKWLDAHE
 orf146ng-1 RHARRIRIDTAINPELEALAELHLYQWQGFWLSTNNRQEISALVILLQTRRKWLDAHE
 orf146-1.pep RQHLRQSLSLTREHGK
 60 orf146ng-1 RQHLRQSLSLTREHGK

Furthermore, ORF146ng-1 shows homology with a hypothetical *E. coli* protein:

sp|P33011|YEAA_ECOLI HYPOTHETICAL 40.0 KD PROTEIN IN COBU-SBMC INTERGENIC REGION
 >gi|1736674|gnl|PID|d1016553 (D90838) ORF_ID:c348#20; similar to [SwissProt
 Accession Number P33011] [Escherichia coli]>gi|1736682|gnl|PID|d1016560 (D90839)
 ORF_ID:c348#20; similar to [SwissProt Accession Number P33011] [Escherichia coli]

```

>gi|1788318 (AE000292) f352; 100% identical to fragment YEEA ECOLI SW: P33011 but
has 203 additional C-terminal residues [Escherichia coli] Length = 352
Score = 109 bits (271), Expect = 2e-23
Identities = 89/347 (25%), Positives = 150/347 (42%), Gaps = 21/347 (6%)
5
Query: 20 YRHRRLLHARVLGGTFLFATALARLLHLHQEWIGMTVEVFLGMLQFQGAIYSNAVERML 79
YRH R++H R+ L + RL + W +T+ V++G + F G + A ER+
Sbjct: 15 YRHRYIVHGTVALAFLLTFLIIIRLFTIESTWPILTMVIMGPISFWGNVVPRAFERIG 74

10
Query: 80 GTVIGLCAGLGVLNWLHGFHGNLTLASALAGWAAVGNGNYYVPMGLAQGATLTCMMLI 139
GTV+G GL L L L + A L GW+A+K Y +L G+T+ +++
Sbjct: 75 GTVIGLSILGLIALQLE---LISLPLMLVWCAAAMELCGWLALGKPKYQGLLIGVTLAIIV 131

15
Query: 140 GDNGSENLDSGLMRAMNVLIHGXXXXXXKLLPQLSTLMWNRMLADNLDACSKMIAEINS 199
G E +D+ L R+ ++V++G + P ++ + W+ LA +L + +++ + +
Sbjct: 132 GSPTGE-IDTALWRSGDVILGSLLAMILPTGIPWQPRAFIHWRNRIQLAKSLTEYNRVYQSAFS 190

Query: 200 GRRMTRERLEONMVVKMRQINARMVKSRSHLAATSGESRISPSMMEAQMHAHRKIVNXXXX 259
+ R RLE ++ K+ VK R+ S+E+RI S+ E +Q +R +V
Sbjct: 191 PNLLERPRLESHQKL--TDAVKMRGLIAPAKSETRIPKSIYECIQTINRNLCMEL 247

20
Query: 260 XXXXXXXXQSPK---LNGSEIRLLDRHFXXXXXXXAALINGRHARRIRIDTAINPEL 316
+ LN ++R D AL G +N +
Sbjct: 248 QINAYWATRSPHSFLNALQKLR--DTAVHMMQOILLSLVHALYEGNPQPVFANTEKLNDAV 305

25
Query: 317 EALAEHLI-HYQWQ-----GFLWLSTNMQRQEISALVILLQRTRK 354
E L + L H+ + G++WL+ ++ L L+ R RK
Sbjct: 306 EELRQLNNHHDLKVETPIYGWVNMETAHQELLSNLICRALRK 352

```

On the basis of this analysis, including the identification of several transmembrane domains in the 30 gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 76

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 639>

```

35      1 ..GCCGAAGRCGA CGCCGGTTAC CGGCACGCGT TTGAGCCCGT ACGGCCATTCA
51      GGGCAAACTC CTGACTGTGC CGCAACACAA CGAACCGCGC ATGGCGGACA
101     AGATTGTCGG CTATCTTCA GACCGCATGG TTGTCGACA GGTTTCCGAT
151     CGGGGTGCGC CGGGCGCTGTG CGACCCGGGC CGGAAACTCG CCCGCGCGT
201     GCCTGAGGCC GGGTTTAAG TCCTTCCCG CCTGGGCGCA AC.GCGGTGA
251     TGCGGCCCTT GAGCGCTGGCC GGTGTGGAAG GATCCGATTT TTATTTCAAC
301     GTTTTTGATC CGCCGAATCT GGGGAGAGCG AGGAAACTGT TTGCCAAATG
351     GTGCGGGCCG GGGTTTCTTA TGCTCATGTG TGAACACCGG CACCGCATGG
401     GTGCAGCGCT TGGCCATATG CGCGAACCTG TCCCGGAAAG CGGATTAATG
451     CTGGCCGCCG AAATCATGGAA ACAGTTGGAAG ACAGTTCTTA CGCGCACCGT
501     TGGGGAATTG CAAGCGGCAT TGCTCGCGA CGCGCACCAA TC CGCGCGCG
551     AGATGGTTT GTGTCTTTAT CGCCGGCAGG ATGAAAACAC CGAAAGCTTG
601     TCCGACTCG CGCAGAACAT CATGAAAATC CTCACAGCG AGCTGCCGAC
651     CAAACAGCGG CGCGAGCTTG CTGCGAAAT CACGGGCGAG GGAAAGAAAG
701     CTTTGACGA T..

```

This corresponds to the amino acid sequence <SEQ ID 640; ORF147>:

```

50      1 ..AEDTRVTAQL LSAYGIQGKL VSVREHNERQ MADKIVGYLS DGMVVAQVSD
51      AGTPAVCDP AKLARVRREA GFKVVPVNGA XAVMAALSA VGEVSDPYFN
101     GFVPFKPSGER RXLFALKVNRFA APFIVMFETP HRIGAALADM AELFPERRLM
151     LAREITKTFI TLFSCTVQEI QTALSADGDQ SRGEMVLVLY PAQDEKHegl
201     SEASONIMKI LTAELPLTKQA AELAALKITGE GKKAlyD..

```

55 Further work revealed the complete nucleotide sequence <SEQ ID 641>:

```

1 ATGTTTCGAGA ARCATTTGCA GAAAGCTCC GRCAGCGCTCG TCGGAGGGAC
51 ATTATCACCTG GTTGCACGCG CATCGGCAA TTGGCGGAC ATTACCCCTGC
101 GCGCTTGGC GGTATTGCAA AAGGCGGACA TCATCTGTG CGAAGACACG
151 CGCGTTACCG CRACGCTTT GAGCGCGTAC GGCATTCAAG GCAAACCTCGT

```

201 CAGTGTCGCG GAACACAACG AACGGCAGAT GGCAGCACAG ATTGTCGGCT
 251 ATCTTCAAGA CGCGCATGGT GTGGCACAGG TTTCGATGC GGTTACGCG
 301 GCGGTGCGC ACCGGCGCGG GAAACTCGGC CGCGCGCTGC GTGAGGCCGG
 351 GTTTAAAGTC GTTCCCGCTG TGCGGCCAAG CGCGGTGATG CGCGCTTGA
 401 GCGTGGCGG TGTGAAAGG TCCGATTTT ATTCAACCG TTTTGACCG
 451 CGGAAATCTG GAGAACGCCA GAAACTGTTT GCCAATGGG TGCGGCCGCG
 501 GTTTCCTATC GTCATTTTC AAACSCCGCG CGGCATCGGT GCACGCGCTG
 551 CGGATATGCG GGAACCTGTC CGCGAACGCC GATTAAATGCT GGCGGCCGAA
 601 ATTACGAAA CGTTGAAAC GTTCTTAAGC GGCAACGGTG GGAAATTC
 651 GACGCCATTG TCTGCCGAGC GCAACCAAATC CGCGGCCGAG ATGTGTTGG
 701 TGCTTTATCC GGCGCAGGT GAAAACACCO AAGGCTTGTG CGAGTCGGCG
 751 CAAACACATCA TGAAAATCTT CAACAGCGGAG CTGCCGACCA AACAGGCCGC
 801 GGAGCTTGCT GCGAAATCA CGGGCGAGG AAAGAAAAGCT TTGTACGATC
 851 TTGGAAAAAC AAATAG

15 This corresponds to the amino acid sequence <SEQ ID 642; ORF147-1>:

1 MFQKHLQKAS DSVVGGTLVY VATPIGNLAD ITLRLAVALQ KADIICAEDT
 51 RVTAAQLSAY GIQGKLVSVR EHNERQMADK IVGYLSDGMV VAQVSDAGTP
 101 AVCDPGAKLA RRVEAGKFLV VFPVGASAVM AALSVAGVEG SDFYFNGFVP
 151 PKSGERKLFK AKWVRRAFFI VMFETPHRIG ATIADMAELF PERRMLARE
 201 IKTTFETFLS GTVGEITQAL SADGNQSRGE MVIVLYPAQD EKH EGLSESA
 251 QNIMKILTAE LPTKQAAELA AKITGEKKKA LYDLALSWN K*

Computer analysis of this amino acid sequence gave the following results:

Homology with hypothetical protein ORF286 of *E.coli* (accession number U18997)

ORF147 and *E.coli* ORF286 protein show 36% aa identity in 237aa overlap:

25	orf147: 1	AEDTRVTAQLLSAYGIQGKLVLVSREHNERQMDAKLKVGYLSDGMVVAQVSDAGTPAVCDPG 60
		AEDTR T LL +GI +L ++ +HNE+Q A+ ++ L +G +A VSDAGTP + DPG
	orf286: 43	AEDTRHTGLLQHFGINARLFLAHNDHNEQQKAETLAKLQEGQNLIALVSDAGTPLINDPG 102
30	orf147: 61	AKLARRVHEXXXXXXXXXXXXXXXXXXXXXXEGSDFYFNGFVPPKSGERRKLFPAWKVRA 120
		R RE F + GE+P K S RR
	orf286: 103	YHLVTRCREAGIRVVPFLPGPCAITALSARGALPSDRFCYGEGLPAKSKGRRDALKAIEAE 162
35	orf147: 121	APFVQMFETPHRIGALADMAELFPEPLLMLAREAITKTFETFLSGTVGEIQTALSADGD 179
		++ +E+ HR + +L D+ + E R + +LARE+TKT+ET VGE+ + D +
	orf286: 163	PRTLIFYESTHRLLSDLEDIVAVLGESERYVLLARELTKTWTETIGHAPVGELLAWVKEDEN 222
40	orf147: 180	QSGRNVLVLYPAQDEKHEGLSESQAQNIMKILTAELPTKQAAELAAKITGECKKALY 236
		+ +GENVL++ E L A + +L ASELP KAAA LAA-I G K ALY
	orf286: 223	RRKGENVLIV-EGHRAQEEDLPLADALRTLALLQAEPLKKAAAELAEIHGVKKNALY 278

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF147 shows 96.6% identity over a 237aa overlap with ORF75a from strain A of *N. meningitidis*:

45	orf147.pep	10 20 30	AEDTRVTAQLLSAYGIQGKLVLVSREHNERO
	orf75a	TLYVVATPIGNLADITLRLAVALVQKADIACEDTRVTAQLLSAYGIQGKLVLVSREHNERQ	
		20 30 40 50 60 70	
50	orf147.pep	40 50 60 70 80 90	MADKIVGYSLSDGMVVAQVSDAGTPAVCDPGAKLARRVREAGFKVVPVGAXAVMAALSAV
	orf75a	MADKIVGYSLSDGMVVAQVSDAGTPAVCDPGAKLARRVREVGFKVVPVGAXAVMAALSAV	
		80 90 100 110 120 130	
55	orf147.pep	100 110 120 130 140 150	GVEGSDFYFNGFPVKSGERRKLFAKWRAAFPIVMEFTPHRIGAALADMAELFPERRLM
	orf75a	GVAGSDFYFNGFPVKSGERRKLFAKWVRVAFPVVMFETPHRIGATLADMAELFPERRLM	
		140 150 160 170 180 190	
60	orf147.pep	160 170 180 190 200 210	LAREITKTFTLSGTVGEIQTALSADGDQSRGENVLVLYPAQDEKHEGLSESAQNIMKI

	orf75a	LAREITKTFETFLSGTVEIQTALADGNQSRGENVLVLYPAQDEKHEGLSESQAQNIMK1	
		200 210 220 230 240 250	
5		220 230	
	orf147.pep	LTAELPTKQAAELAAKITGEGKKALYD	
	orf75a	LTAEPLITKQAAELAAKITGEGKKALYDLALSWKNX	
		260 270 280 290	

10 ORF147a is identical to ORF75a, which includes aa 56-292 of ORF75.

Homology with a predicted ORF from *N.gonorrhoeae*

ORF147 shows 94.1% identity over a 237aa overlap with a predicted ORF (ORF147ng) from *N. gonorrhoeae*:

15	orf147.pep	AEDTRVTAQLSSAYGIQGKLVSREHNERQ	30
	orf147ng	TLYVVATPIGNLADITLRLAVLQKADIICAEDTRVTAQLSSAYGIQGKLVSREHNERQ	85
	orf147.pep	MADKIVGYLSGDMVVAQVSDAGTPAVCDPGAKLARRVRREAGFKVVPVVGAXAVMAALSVA	90
20	orf147ng	MADKIVGFLSGLVVAQVSDAGTPAVCDPGAKLARRVRREAGFKVVPVVGASAVMAALSVA	145
	orf147.pep	GVEGSDFYFNGFVPPKSGERRKLFAKFWRAAFPIVFMETPHRIGAALADMAELFERRLM	150
	orf147ng	GVAEASDFYFNGFVPPKSGERRKLFAKFWRAAFVVMETPHRIGATLADMAELFERRLM	205
25	orf147.pep	LAREITKTFETFLSGTVEIQTALADGSQSRGENVLVLYPAQDEKHEGLSESQAQNIMK1	210
	orf147ng	LAREITKTFETFLSGTVEIQTALADGNQSRGENVLVLYPAQDEKHEGLSESQAQNAMK1	265
30	orf147.pep	LTAELPTKQAAELAAKITGEGKKALYD	237
	orf147ng	LAAELPTKQAAELAAKITGEGKKALYDLALSWKNX	300

An ORF147ng nucleotide sequence <SEQ ID 643> was predicted to encode a protein having amino acid sequence <SEQ ID 644>:

35	1	MSVTOAFFPM FQHHLQKASD SSVGGTLYVVV ATPIGNLADI TIRALAVLQK	
	51	ADIACEDTR VTAQQLSSAYG IQGRLVSVRE HNERQMDKV IGFSLSDGLVV	
	101	AVQVSDAGTPA VCDPGAKLAR RVREAGFKVVPVVGASAVMA ALSVAGVAGAES	
	151	DFYFNGVVPK KSGERRKLFA KWVRAAFFVVM METPHRIGA TLADMAELF	
	201	ERRMLAREI TKTTFETFLSGTVEIQTALADGNQSRGENVLVLYPAQDEKHEGLSESQAQNIMK1	
40	251	KHEGLSESQAQ NAMKILKAAEL PTKQAAELA KITGEGKKAL YDLALSWKNX	
	301	*	

Further work revealed the following gonococcal DNA sequence <SEQ ID 645>:

45	1	ATGTTTCAGA AACACTTGC GAAAGCCTCG GACAGCGTCG TCGGAGGGAC	
	51	ATTATTCAGTG GTTGCACCGC CCATCGGCCA TTGGCGCACG ATTACCCCTGC	
	101	GCGCTTTGGC GGTTATTCGA TCACTGGCGA CCGAGACAGC TCATTTGTC CGAAGACAGC	
	151	CGCGTTACTCG CGCAGCTTTT GAGGGCGTAC GGCATTCAGG CGAGGTGGT	
	201	CGACTGTGGC GAAACAAACG AGCGCGAGT GCGGACACAG GTAACTCGTT	
	251	TCCCTTCAGA CGGGCTGGTT GTGGCGCAGG TTTCGATGC GGTRACCGCG	
	301	GCGCTGTGGC ACCCGGGCGC GAAACTCGCC CGCCCGGTGCG CGAACGAGG	
	351	GTTCAAAAGTC GTTCCCCTCG TGCGCGCAAG CGCGGTTAATG CGCGCGTTGA	
	401	GTGTCGGCGC TGTCGGCGAA TCAGGTTTTT ATTCAAACGG TTGGTACCG	
	451	CGGAATTCGCG CGAACGACTCG GAAATTGTTT GCGAAATGGG TCGGGCGGC	
	501	ATTCCTGCTG GTCATTTGTA AAACCGGGCA CGGAATCGGG CGAACGCTTG	
55	551	CGGATATGGC GGAAATTGTC CCCGAACGCCG GTCTGATGTC CGCCGCCGGA	
	601	ATCCGAATTA CGTTGAACG GTTCTTAAGC GGCACTGGTTS GGAAATTCGA	
	651	GACGGCATTC GCGCGGCGAC GCAACCAATC CGCCGCCGAG ATGGTGTGCG	
	701	TGCTTATTCG CGCGCGGAGT GAAAACACG AAGGCTTGTG CGAGTCTGGC	
	751	CAAATATGCGA TGAATATCT CGTGGCGGAG CTGGCGACCA AGCACGGGGC	
	801	GGAGCTTGGC GCCAAGATTA CAGGTGAGGG CAAAAAGGCT TTGTACCGATT	
60	851	TGGCACTGTC GTGGAAAAAC AAATGA	

This corresponds to the amino acid sequence <SEQ ID 646; ORF147ng-1>:

```

      1 MFQKHLQKAS DSVVGGTLVV VATPIGNLAD ITLRLALAVLQ KADIICAEADT
      5 RVTAQILLSAY GIQGRIVSVR EHNERQMADK VIGFLSDGIV VAQVSDAGTP
     10 AVCDPGAKLA RRVREAGFKV VPVVGSAVM AILSVAGVAE SDFYFNGFVP
     15 PKSGERRKLK AKWVRAAFTPV VMFETPHRIG ATLASMAELF PERRILMLARE
     20 ITKTFTFPLS GTVGEIQTAL AADQNQSRGE MVLVLYPAQD EKHEGLSESA
     25 QNAMKILAAE LPTKGAAEAL AKITGEGKKA LYDIALSWKN K*

```

ORF147ng shows homology to a hypothetical *E.coli* protein:

```

10 sp|P45528|YRAL_ECOLI HYPOTHETICAL 31.3 KD PROTEIN IN AGAI-MTR INTERGENIC REGION
(F286)
>gi|606086 (U18997) ORF_f286 [Escherichia coli]
>gi|1789535 (AE000395) hypothetical 31.3 kd protein in agai-mtr intergenic region
[Escherichia coli] Length = 286
Score = 218 bits (550), Expect = 3e-56
Identities = 128/284 (45%), Positives = 171/284 (60%), Gaps = 4/284 (1%)
Query: 4 KHLQKASDVSVGGTLVV VATPIGNLADITLRLALAVLQKADIICAEADT
      5 R K O A +S G LY+T TPIGNLADIT RAL VLQ D+I AEDTR T LL +G+
Sbjct: 2 KQHQSAQNSQ--GQIYIVPFTPIGNLADITQRQALEVQAVDILIAEADTRHTGLLQHNM 59
20 Query: 64 GRLVSVEHRRQMDKAVIGFLSLDGVVQAQVSDAGTPAVCDPGAKLARRVREAGFKVPPV 123
      5 RL ++ +HNE+Q A++ + L +G +A VSDAGTP + DPG L R REAG +VWP-
Sbjct: 60 ARLFALHDHNEQQKAETLIAKLOEQEQNIALVSDAGTPLINDPGYHLLVTRCREAGIRVVPL 119
25 Query: 124 VGASVNAVMAALSVAGVAESDFYFNGFVPPKSGERRKLFAKWKVRAAFTPVVMFETPHRIGATL 183
      5 G A + ALS AG+ F + GF+ KS RR + + +E+ HR+ +L
Sbjct: 120 PGPCNAITALSAAGLPLSDRFCYEGFLPKSKGRDLAKIAEAPRTLIFYESTHRLLDSDL 179
30 Query: 184 ADMAELPFEER-LMLAREIITKTFETFLSGTVEIQTALAADGNOSRGENMLVLYPAQDEK 242
      5 D+ + E R ++LAE-E-TKT+E+ VGE+ + D N++ +GEMVL++ +
Sbjct: 180 EDIVAVLGESRYVVLARELTKTWEIHPGAPVCELLAWVKEDENRKGEMVLIV-EGHKAQ 238
35 Query: 243 HEGLESAAQNAMKILAAEPLTQKQAAELJAKITGEKKALYDIAL 286
      5 E L A + + LELP K+AA LAA+I G K ALY AL
Sbjct: 239 EEDLPADALRTLIAILQAEPLPKKAALAAEIHGVKKNALYKYL 282

```

Based on the computer analysis and the presence of a putative transmembrane domain in the gonococcal protein, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 77

40 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 647>

```

      1 ATGAAAACAA CCGACAAACG GACAACCGGA ACACACCGCA AAGCCCCGAA
      5 AACCGCTGCG ATCCGCTCT C. GCTCTTTA CTTAGCCATA TGCCCTGTGTT
     10 TCGGCATTC TCCCCAACGC TGGCCGGGA ACACATTATT CGGCATCACAC
     15 TACCAAATACT ATCGGCACAT TGGCCAAAAA AAAGGCAAGT TTGCACTGG
     20 GGCGGAAGAT ATTGAGTTT ACAACAAAAA AGGGGAGTTG TTGGCAAAAT
     25 CAATGACAAA AGCCCCCATG ATTGATTTT CTGTGGTGTC CGCTAACGCC
     30 GTGGCGGCAAT TGGTGGCGC: ATCAAATAT TGTGAGGCTG GCACAAACG
     35 GCGGCTATAA CAACTTGTAT TTACTTATAA ATTTGTAAGA CGGAATAATT ATAAGGCAG
     40 CAACAGCGW TTACTTATAA ATTTGTAAGA CGGAATAATT ATAAGGCAG
     45 GACTTAAAGGC CRATCTTATG GCGGGCATTA TCATATGCC CGTTGCTATA
     50 AATWTCTCAC AGATGCGAGA CCTGTTGAA TGACCAGTTA TATGGATGGG
     55 CGGAATATA TCGCAATTA TAATACCCCT GACCGTGTTC GTATTGGGC
     60 AGGCAGGGCA TATTGGCGAT CTGATGAAAG TGACCCCAAT AACCGGGAAA
     65 GTTCAATCA TATTCGAACT ..... .
    701 ..... .GCTC ACCAAATGTTT ATCTATGATG CCCCAGAA
    751 AAAGCTGGTTA ATTAATGGG CCTTGCAAC GGGCAACCCC TATATAGGAA
    801 AAAAGCAATGG CTTCCACCTG GTTGTAAAG ATTGGTCTTA TGATGAAATC
    851 TTTCGCTGGAG ATACCCCATTC AGTATTCCTAC GAACCCAGTC AAAATGGGA
    901 ATACTCTTT AACGACGATA ATAAATGGCAC AGGAAATAC AATGCCAAAC

```

5	951	ATGAAACCAA CTTCTCGICG ATAAAGTAA ACCTAACGAC GCTTAATGTT 1001 TTTAATGTTT CTTCATTCGA CGACCGACCA GAACCTGGT ATCCATGCTGG 1051 AGGTGCTGTC AACAGCTATC GACCCGACT GATAATGGA GAAAATTTG 1101 CCTTTATGCA CGGAAAAAA GGCGATGTA TACTTACCG CANCATCAAT 1151 CAANGGTGTC GAGGATTATA TTTCAGGA GATTITTAACGG TCTGCCTG 1201 AAATAACGAAC ACTTGGCA CGCGCCGGCT TGATCATGAG GAGGAGCATG 1251 CCTCTTACTG GAAAGAACG GGGTGGGCCA ACAGCGGCCGTC GCAAAATTC 1301 GGCAGAACCA CGCTG.....	..	
10	2101	2101GATAAAG	
2151 TGACGTCCTC ATTGACTAAG ACCGACATCA GCGGCATGTC CGATCTTGCC 2201 GAT/GCAGCTC ATTAAATCTC CACAGGGCTT GCGCACATCA CGACGCAATC 2251 TAGTCGCAATC GGCATACAC TTGATTAACGT GCGCACACAC GCGCAACAA 2301 ACGCACACCC TAGCTCCG G. SANTGCGC AAGCGAACAT GATTCCTTCAT 2351 ATCAATTAACG GCAACACATC GGGCTTCGGC ATGACTTCAT TTATACCTAAG 2401 CGACCCACCC TGATCAACG CCGACTTCGCA CGACTGTCGAC CTTCCTGGG AACATAACG 2451 CAAAGCTTACG CCATTCTGGCA CTCAACCGDTA ATGTTCTCCC AGCGCAATAAG 2501 CGAGATTCAC TTTTGTGAA CAGCGCCCTT ATCCGGAAACAA CGGAGGGCGG 2551 Ga:CGGAATTG GAACTTACAT TAAAGAACGAC CGAATGTCGAG CTGGCGCTCAG 2601 Ga:CGGAATTG GAACTTACAT AACCCTTGACA CGGCACCATC ATCACTCAAA 2651 TCCGGCTATC GCCAGCATGC CGGAGGGGGG AACAGGGCGG TGCGGAGACAG 2701 TGGGGCCGGC CGCCGCTTGC CGCTTGTGGG CGCTTCTCTTA TTATCTTCA 2751 CACCCGACAC TTCGGTAGAA TCCCTTTCA CGACGCTGCG ATGTTACAGCG 2801 AAATTTGAAAC GTGCCGGACG ATTCGGCTTT ATGTCGGAGA TTCTGGCTCA 2851 CGCCAGGAGC AAATTGAAAC TTGGCGGAAG CGGAGAACGG ACTTACACCT 2901 TGCCGGTCAA CAATPACCGG AACGACCTCG CAAGCTCTGA RCAAATTGAGC 2951 GTAGTGAGAG GAAARAGCA CAAACCGCTG CCGGAAACAC TTAAATTTCAC 3001 CCTGCAAAAC GAACTCGTG ATTCAGGGCGG GTGG.....	..			
30	3551	3551TTAGAC CGCGTATTTC CGGAGAACCG 3601 CGCGAACGCC GTTTGGACAA CGGGCATCTCG CGAACACAAAG CACTACGGCT 3651 CGCAAGATTT CGGGCCTCA CGCCACAAAG CGGACCTGGT CGAACCGGT 3701 ATCGAGAAAAA ACCCTCGGGC CGGGGGGGG GCGCATCTGT TTGGTCACAA 3751 CGCGAACGCC AAACCTTCGG CGAACGGCAT CGGCACACTG CGACGGCTTGG 3801 CCCAGCGGCC CGTTTTCGGG CAATACGGCA TGACAGCGG ATCTACGGCT 3851 ATCAGCGGCC CGGGCGGGGT TTAGACGGG GACCGCTTCA GACGGCTATCG 3901 GAGsMmAAtt CGCGGCCGGT GTGTCGCTATT CGGCATCTGA CGGACGATAC 3951 CGGGCGGCT GtttCGGATT CGGCATCTGA CGGCACATCG CGGCACAGCG 4001 cTTATTTGCTC AAAAAGCGC ATTACGGCTA CGAAAGCTG ATTAACGGCA 4051 CGGGCGGCTT TCATTCGAAG CGCCTACCGG CGGCATCTGA CGGACGATAT 4101 TCTATTCAAC CGGGCGGCA CATTTCGCTAC CGGCCTTCTG TGACGGCTCTG 4151 CTATCCATGG CGGCCCTTCGA CGAAAGCTCG AACAGCGCTG ATAACCGCG 4201 TATTGGCTCA GGATTTCGGC AAAACCGCGA STGCGGAAGT GggCTAAAG 4251 CGGAAATCA AAAGTTTTCAC GGTCTGGCTC CGACGGCTCG CGGGCGGAAAG 4301 CGCGAACACTG GAAGCGCAAC CGACGGCGGG CATCAAAATTA GGCTACCGCT 4351 GTGAA.....	..
35	3601 CGCGAACGCC GTTTGGACAA CGGGCATCTCG CGAACACAAAG CACTACGGCT 3651 CGCAAGATTT CGGGCCTCA CGCCACAAAG CGGACCTGGT CGAACCGGT 3701 ATCGAGAAAAA ACCCTCGGGC CGGGGGGG GCGCATCTGT TTGGTCACAA 3751 CGCGAACGCC AAACCTTCGG CGAACGGCAT CGGCACACTG CGACGGCTTGG 3801 CCCAGCGGCC CGTTTTCGGG CAATACGGCA TGACAGCGG ATCTACGGCT 3851 ATCAGCGGCC CGGGCGGGGT TTAGACGGG GACCGCTTCA GACGGCTATCG 3901 GAGsMmAAtt CGCGGCCGGT GTGTCGCTATT CGGCATCTGA CGGACGATAC 3951 CGGGCGGCT GtttCGGATT CGGCATCTGA CGGCACATCG CGGCACAGCG 4001 cTTATTTGCTC AAAAAGCGC ATTACGGCTA CGAAAGCTG ATTAACGGCA 4051 CGGGCGGCTT TCATTCGAAG CGCCTACCGG CGGCATCTGA CGGACGATAT 4101 TCTATTCAAC CGGGCGGCA CATTTCGCTAC CGGCCTTCTG TGACGGCTCTG 4151 CTATCCATGG CGGCCCTTCGA CGAAAGCTCG AACAGCGCTG ATAACCGCG 4201 TATTGGCTCA GGATTTCGGC AAAACCGCGA STGCGGAAGT GggCTAAAG 4251 CGGAAATCA AAAGTTTTCAC GGTCTGGCTC CGACGGCTCG CGGGCGGAAAG 4301 CGCGAACACTG GAAGCGCAAC CGACGGCGGG CATCAAAATTA GGCTACCGCT 4351 GTGAA.....	..		
40	4001 cTTATTTGCTC AAAAAGCGC ATTACGGCTA CGAAAGCTG ATTAACGGCA 4051 CGGGCGGCTT TCATTCGAAG CGCCTACCGG CGGCATCTGA CGGACGATAT 4101 TCTATTCAAC CGGGCGGCA CATTTCGCTAC CGGCCTTCTG TGACGGCTCTG 4151 CTATCCATGG CGGCCCTTCGA CGAAAGCTCG AACAGCGCTG ATAACCGCG 4201 TATTGGCTCA GGATTTCGGC AAAACCGCGA STGCGGAAGT GggCTAAAG 4251 CGGAAATCA AAAGTTTTCAC GGTCTGGCTC CGACGGCTCG CGGGCGGAAAG 4301 CGCGAACACTG GAAGCGCAAC CGACGGCGGG CATCAAAATTA GGCTACCGCT 4351 GTGAA.....	..		
45	4351 GTGAA.....	4351 GTGAA.....	..	

This corresponds to the amino acid sequence <SEQ ID 648: QRE1>

50	<p>1 MKTDTKRTTE THRKPKTGTR IRFXAYLAI CLSFGLPQA WAGHTYFGIN 51 YQYDFRDAEN KGFVAGKAD LEVNGXKGEL VKGSMKAPM IDFSVSRM 101 VAALVGVOYI VSVAHNGGIN VNDFGAEGK EXDQXRTXYK IVKRNHYKAG 151 TKGHPYGGDY HMPRLHLKXV DAEPEMTSY MDGRYIDQK NYPRVRIGA 201 GROYWRSDSE EPNNRESSYH ISAS..... GS PMFTYDQOK 251 KWLLNLVQLQ GPNYPIGKSQF FOLVRNGVLT DEIFAGDTHS EPPYERQNK 301 YSFNDNNNNTG GKNIAKEHG SPBLNRLKTTR VOLFNVLSE TAREPVYHA 351 GGVNSYRPL NNGENISFD EGKGLHNGT NIINGAGGGLY FOGDFTVSE 401 NNETWQAGV HISEDSTVTW KVNGVANRDL SKIGKTL // </p>
55	<p>701 DRKVTA S LTKTDISGNV LDADAHHLNL TGLATLNGNL 751 SANGDRTYRV SHNATONGN SLVXNAQAF NQATLNGNS ASQGNAFLS 801 DHAVQNSLTL LSGNAKANVS HGAINGVNSL ACDAHVFES SRFTQISGG 851 KDTALHLKDS EWTLPGCF GLHNLDNATI TLNASYRHD AAGCTGSATD 901 APPRRRSRSL RSLXLXPTP SEVSRFNLT QVNKLNGGCT PLSFLFEGY 951 RSDKLKLAE SEGTYYFLAVN NTGNEPASLE QLTWVEGKDQ KPLESENINF 1001 LQNHEVDAGA W // </p>
60	<p>1151 LDRWFVAEDR 1201 RNAVNGSCIN DTKHNRQSDF RAYRQFTDLR QIGMGNLGS GRVGILSHP 1251 RTENTFDGDI GNSARLAHG FVGOGYIDFR YIGISAGAG SSSGLSDCG 1301 XKXRRLVLY GIQARYRGF GFGFELPHIG ATRYFVVKAD YRYEYNVNTI 1351 PLCAFNRVYR GIKADYSFKP AHOISITYP SLSYTDASGA KVTRVNTAV</p>
65	
70	

1401 LAQDFGKTRS AEWGVNAEIK GFTLSLHAAA AKGPQLEAQH SAGIKLGYRW
 1451 *

Further sequencing analysis revealed the complete nucleotide sequence <SEQ ID 649>:

5 1 ATGAAACACAA CGGACAACAGC GACAACCGGA ACACACCGCA AAGCCCGAA
 51 AACGGCCGC ATCCGCTTCT CGCGCTTCA CTTAGCCATA TGCCCTGCTG
 1001 TCGGATCTCT TCCCGAACGCC TTGGCCGGAC ACACATTATT CGCGCATCAC
 151 TACCAACT CTGGCGACT TGCGGAAAT AAAAGCAAGT TTGCGATCGG
 201 GGCGAAAGAT ATTGAGGTTT ACACAAAAA AGGGGAGTTG GTGCCGAAAT
 251 CANTGACAAA AGCCCCGATTA ATIGATTTTTG CTGTTGGTGC GCGTAAACGG
 301 GTGGCCGCACT TTGTGGGCCA TCACATATTAT TTGAGCGCTGG CACATAAACGG
 351 CGCGCTTACAC AACCTGGTAT TTGGCCCGAA AGCGAGAAAT CCCGATCAC
 401 ATCGGTTTAC TTATTAATATT GTGAAACGGA ATAATTATA AGCAGGGACT
 451 AAAGGCCATC CTATGGCGG CGATTTATCAT ATGCGCGGT TTGCTATAATT
 501 TGTACACAGAT GCAGAACCTG TTGAAATGAC CAGTTATATG GATGGCCGGA
 551 AAATATGAGA TCAAAATATA TACCCCTGAC TTGTTGCTGTAT TTGGGCAAGGC
 601 AGGCAATTATG GCGGATCTGA TGAGATGAC CCCAAATAACG GCGAAAGITC
 651 ATATCAATTG CAAGGTGCGT ATTCTTGGGC CTTGGTGTGAC ATAACCTTGT
 701 CAAACATTGG ATCAGGTGGT GGCAGACGCA ACTTAGGTAG TGGAAAAATT
 751 AAACATAGCC CATATGGTTT TTGACCAACG GAAGGGCTCAT TTGGCGACAG
 801 TTGGCTACCA ATGTTTATCT ATGATGCCA AAAGCAAAG TTGTTAAATTA
 851 ATGGGGTTAT GCAAAACGGG AACCCCTATA TAGGAAAAAG CAATGGCTTC
 901 CAGCTGGTTC GTAAAGATG GTTCTATGAT GAAATCTTGG CTGGAGATAC
 951 CCATTCTAGTA TTCTACGAC CACGGTAAAGA TTGGGAAATAC TCTTTAAACG
 1001 AGCATTAATTAACG CACGGACGAA AAAATCAATG CAAACATGAA ACACAACTCT
 1051 CTGCTTAATTA GATTTAAACG AGGAACGCTT CAATTGTTTA ATGTTTCTTT
 1101 ATTCGAGACA GCAAGAGAC CTTGTTATCA TGCTGCAAGT GTGTCGAAACA
 1151 GTTAAAGGCG CAGACGTAAAT ATGGAGAAA ATATTCTCTT TATTGACCAA
 1201 GGGAAAGGCG AATZGATACT TACCAACAC ATCAATCAAG GTGCTGAGGG
 1251 ATTATTTTC CAAGGGAGT TTACGGCTTC GCCTGAAATAC ACGAAATCTT
 1301 GGAACGGCG GGGCTCTCAT ATCACTGAG ACAGTACCGT TACTTGGAAA
 1351 TAAACAGCGC TGCAAAAGCA CCACCTGTCC AAAATCGGA AAGGCACCT
 1401 GCACGTTCAAC GCAAAAGGGG AACAAACAGG CTTCGATCAGC GTGGGCACG
 1451 TGTACAGCTT ATTCGATGCA CAGGCAGACO ATAAAGGCAA AAAAACAGG
 1501 TTTATGGAAA TGCGCTTGTG CAGCGCGGAGG GOTACGGTGC ACTGAATGCG
 1551 CGTATTAATCG TCAACCCCTA ACAACATTA TTTCGCTTGT CGGGCGGAC
 1601 GTTGGGTTATT AAACGGGCAT TCGCTTCTGCT TCCACCGTAT TCAAAATACCG
 1651 GATGAAGGGG CGATGATGTT CAACCCAAAT CAAGACAAAG ATCCACCGT
 1701 TACCATACCA GCGCAATTAAG ATATTCTGAC AACCCGCAAT AACAACAGCT
 1751 TTGATGACAA AAAAGAAATG GCTCATACCO TTGTTGTTGG CGAGAAAGAT
 1801 ACCACAAAAAA CGGCGGGGGG GCTCAACCTT TTGTTACCGAC CGCCCGCAGA
 1851 AGACCGSACC CTGCTGCTT CCGCCGAAAC AAAATTAACG GCGAACATCA
 1901 CGCAAAACAA CGCGAACACT TTTTTCAGCA CGAGACCCAC ACGCCAACCGC
 1951 TCAATCATTA TAAAGGCCCA TTGGTCGCAA AAAGAGGGCA TTCTCGGGG
 2001 GGAAATCTCG TGCGAACACG ACTGGATCAAC CGCGCACATT AAAGCGGAAA
 2051 ACTTCACCAAT TAAAGGCCCA CAGGGCGTGG TTTCGGCCA TTGTCGAAA
 2101 GTGAAAGGGG ATTCTGGCTG GACGCAATCAG GCCCAAGCG TTTTGGTGT
 2151 CCACCGCGAT CAAAGGCCCA CANTCTGTAC ACGTTGCGAC TGGACGGTCA
 2201 TGACAAATTG TGCGGAAAAA AGCATTAACCG AGCATTAAGT GATTCCTTCA
 2251 TTGATGACAA CGCGACATCG CGCGCAATGTC GATCTTGCG ATACGGCTCA
 2301 TTTAAATCTC AGAACGGCTTG CCAACACTAA CGGCAATCTT AGTGCACAAATG
 2351 CGCGATCACCG TTATACAGTC ACTGGATCAAC CGGCCAACCTT CGGCCAACCTT
 2401 AGCGCTGCG GCAATGCCCA AGCAACATTAA ATCAAGGCC CATTAAACGG
 2451 CACACATCCG TTGGCTGGCA AGCTCTTATTA TAATCTAACG GACCGACGGCG
 2501 TACAAACAGG CAGTCAGGAG CTTTCGGCCA CGCGTAAGGC AACCGTAAGC
 2551 CATTCCGAC TCAACGGTTA TGTCCTCTTA CGCGATAAAGG CAGTATTCTCA
 2601 TTTTGAAAGC AGCGCTTCA CGCGACAAAT CGCGGGCGG AACGATACGG
 2651 CATTCACTT AAAAGACACG GATGGACCG TGCGGTGAGG CAGGGAAATTA
 2701 GGCAATTAAAC ACCTGTGACA CGCGCACATTAA ACACATCAAT CGCCGCTATCG
 2751 CCACGATCGC CGAGGGGGCA AAACCGCGCA CGCGCACAGAT CGCCGGCGCC
 2801 GCGGTGCGC CGGTGCGCA CGTTCCTCAT TATCGGTAC ACGCCAACAT
 2851 TCGGTAGAAT CGCTTCTGCA CGCGCTGAGC GTAAACGGCA ATTTGAACGG
 2901 TGCGGAAAC TTCCGCTTCA TGTCGGAACT TTGCGGTAC CGCAGGCA
 2951 AATTGAGCTT GCGGGAAAGT CGCGGAAAGCA CTTRACACCTT GGCGGTCAAC
 3001 ATTAACCGCA AGCGACCTTC AGCCCTCGAA CAATTGACCG TATGCGAAGG
 3051 AAAAGACAAAC AAACCGCTGT CGCGAAACCTT TAATTCACCG CTGCAAAACG
 3101 AACAGCTGCA TGCGGGGGG TGCGGTACCG ACCTCATCGG CAAGACCGCG
 3151 GAGTTCGCGC TGCAATTAAT CGTCAGAAAGA CAAGACGTTT CGCACAAACT
 3201 CGGCCAAAGCA GAAGCCAAA AACAGCGGGA AAAAGACAAAC GCGCAAAAGGC
 3251 TTGAGCGCTT GATTCGGCGG CGGGCGGCGT CGCGTACAAAGA GACAGAAAGC
 3301 GTTGGCCGAC CGGGCGGCGA CGCAGCGGGG GAAATGTGCG CGATTATGCA

This corresponds to the amino acid sequence <SEQ ID 650; ORF1-1>;

1 MTKKDTKRTTE THRKAFTKTRT IFRSPYALAI CLSFGLLPLQA WAGHTYHFGIN
 51 YQQYRDFAEN KCGKFVAGARD IEVYNNKGKL VGSKMTKAFM IDFSUVSRN
 101 VAALVGDFYI VSVAHHGNYN NUDVGAEGRN FDGRHPRTKYI VRKNRNYKAGT
 151 KGHPHYGGDH MFRLHFKWTFD AEEFVPYQHNGD YDKRYIDQNI FRTVRIGRA
 201 RQWYRSRDESD PNNKRESEHHYI ASAYSWLVGG NTFAQNQSGE GTVNWLGESEK
 251 KHSFYGFLPT GGSFGSDGSD MFYIADQKQK WLINGVLQTC FTYIGNSG
 301 QLRKWDWUH YELFGATHVS FTYEPYRPNL SFDNDNTNG KINAKHEHNS
 351 LPNRKLTTRV CLFNVLSLNT AREPYVIAHAG GVNSYRPRLG NMENISIPDE
 401 GKGELLITSH INOQAGAGLYF QGDFTVSPEN NETWQAGGVN ISESTDSTWVK
 451 VNGVNLRSN KIGKGTLHRYK AGKENGQHNS VSGDGTILDO ADQGDKKQKA
 501 FSEIIGLVSGH GTVOLNADQN FNFDKLYGF GRRGLDNGL SLSFHPR
 551 DEGAMIVHNH QDKESTVAT GNKHDIAATTGN NNSLSDKKEI VNGWFGKE
 601 TTCTKNGRLN VYQOAEADERT LLLSGCTNLN GNTCTCNGL FSSGRPRTPH
 651 YNHNLHDHSQ KEGIFPRGEV WNDNWINRTE KAELNFOIKG GVSNRWNRA
 701 VKGDNHSLN HSAQEVGAHE QSHCTTRSD WTGLTNCKWE TITDDKVJAS
 751 LTKTIDSGVN DLADHAHNM TGLATLNGNL SANGDTRYT SHNATONGNL
 801 SLVNGQADTF NQATLNGNTS ASGNASNFLS DVHQANGLT LSGNAKANVS
 851 HSALGNVSLN ADEKAVPHFE SRFTGQISGG KOTAHLLKDS EWTLPSTEL
 901 GNLLNLNNTI LTSNAYSRHQA AGACGTSLSAID APPRRRSRRSL RSLSLPVT
 951 SVESRFNTLT VNGKLQGQT FRFMSLQGY RSDKLKLAE SEGTYYLAVN
 1001 NTGNEPASLE CLTVEVGKHN KPLSLENLNFT IONEHWDGAS WRYLQKIRD
 1051 EFRHLNPWKCE CELSDLKLGK AEKQAKAEKD NSALDALTA GRDAVENTES
 1101 VAAEPRAQGG ENVGIMQACE EKKRQVQADKD TALAKOREAT TRATAFFR
 1151 ARRARRDLQO LPQOPQFPQO RDLSIRYANS GLSEFSATL SFVAFQDDEL
 1201 RVFAEDRNRN VWTSGIRDTK HYRSQDFRAY RQQTDLRQIG MOKNGLSSRG
 1251 GFLHSNNTN NTFDGIGNS ARLAANGAVFG QGYDIFRIG YISAGAGFSGSS
 1301 SLDGIGGKIK RNRVLYHKGQI AYRARGFQGF GHEIGYTR VFWDKADYR
 1351 ENVNIAITPLG ATNRYRAGIK ADYSPKPAOH ISITPYLSS YTDASACKVR
 1401 TRVNTWLAQ DFGKTRSAEW GVNAEIKGFT LSLHAAAKG PQLEAQHSAG
 1451 IKLGYWR

Computer analysis of these sequences gave the following results:

Homology with a predicted ORF from *N. meningitidis* (strain A)

55 ORF1 shows 57.8% identity over a 1456aa overlap with an ORF (ORF1a) from strain A of *N. meningitidis*:

	orf1.pep	orf1a	orf1.pep
	MKTTDKRTTETRHKAPTKGIRFXXAYALACILSFGILPQWAHGHTYFGINYQQYRDFAE	MKTTDKRTTETRHKAPTKGIRFSPAYLAICLISFGILPQWAHGHTYFGINYQQYRDFAE	
	10 20 30 40 50 60	10 20 30 40 50 60	
60	orf1a		
			KGKFAVYGAKOIEVYNKKELVGKSMTKAPMIDFSVSVRNGVAALVGVQIVSVASHNGGYN
			70 80 90 100 110
			120
	orf1.pep		

	orfla	KGKFAVGAKDIEVYNKKGEI VGKSMTKAPMIDFSVVSRN GAALVG DQYIVSVARHNGY	70	80	90	100	110	120
5	orf1.pep	NVD FGAEGXNIXDQKRXTYKIVKRNNYAGT KGHPYGGDYIHMPLR HKXVTDAEPEVMTSY	130	140	150	160	170	180
	orfla	NVDFGAEGGXN-PDQHRSFYQIVKRNNYKPDNS-HPYNGDXIHMPLR HKFVTDAEPEVMTSD	130	140	150	160	170	
10	orf1.pep	MDGRKYIDQNNYPDRVRIGAGRQYWRSDEDP-----NN-----	190	200	210			
	orfla	MRGNNTYSDEKEKYPERVIRIGSGHHYINRYDDDKHGDLSYSGAWLIGGNTHMCGWGNNGVXSL	180	190	200	210	220	230
15	orf1.pep	RESSYH---IA----SGSPMFYIDAGKQWKWLNGVLQTGPNPYIGKSNGFQLVRK	220	230	240	250	260	
	orfla	SGDVRHANDYCPMPPIAGAQADSGSPMFYDWTNNKWLLNGVLQTGPPYSGRENGFQLRK	240	250	260	270	280	290
20	orf1.pep	DWFYDEIFAGDTHSVYEPQRNGKYSFNDDNNGTGKINAKHEHNSLPNRLKRTVQLFNV	270	280	290	300	310	320
	orfla	DWFYDDIYRGDTHTVXFEPRSNCHSFTSNNGTGTVTETNEKVSNP-KLKVQTVRLFDE	300	310	320	330	340	350
25	orf1.pep	SLSETAREPVYHAAGGVNSYRPRLNNGENISFIDEKGKELILTSNINQGAGGLYFQGDFT	330	340	350	360	370	380
	orfla	SLINETDKEPVY-AAGGVNQYRPRLNNGENISFIDYGNGKILSNNINQGAGGLYFEGDFT	360	370	380	390	400	410
30	orf1.pep	VSPPENNETWQGACVHISEDOSTTWKVN G VANDRLSKIGKTL-----	390	400	410	420	430	
	orfla	VSPPENNETWQGAGVHISEDOSTTWKVN G VANDRLSKIGKTLHVQAKGENQGSISVGDG	420	430	440	450	460	470
35	orf1.pep	-----						
	orfla	VILDOQQADDKGKKQAFSEIGLXSGRGTVQLNADNQFNFDKLYFGFRGGRLDLNGHSLSFH	480	490	500	510	520	530
40	orf1.pep	-----						
	orfla	-----						
45	orf1.pep	-----						
	orfla	RIQNTDDEGAMIXXHNATTSTVTITGNESITOPSGKNINR NYSKEIAYNGWFGEKDTTK	540	550	560	570	580	590
50	orf1.pep	-----						
	orfla	TNGRINLVYQPAADRTXLSSGGTNLN NGNITQTN GKLFSGRPTPHAYNH LGSGWSKMEG	600	610	620	630	640	650
55	orf1.pep	-----						
	orfla	IPQGEIIVWDNDWIXRTFKAEFH I QGGQAVISRNVAKEVGDXHLSNHAQAVFGVPAHQSH	660	670	680	690	700	710
60	orf1.pep	-----						
	orfla	TICTRSDWTGLTNVCXXITDDKVIASLT KTDXSGVX LXXX XGXAXLXGNL SAN	720	730	740	750	760	770
65	orf1.pep	-----						
	orfla	XXXXXDKVTASLT KTDISGNVLDADH AHLNLTGLATLNGNL SAN	440	450	460	470	480	
70	orf1.pep	GDTRYTVSHNATQNGNXSLVXNAQATFNCQATLNGNTSASGNASFNLSDHAVQNGLSTLSC	490	500	510	520	530	540

-363-

	orf1a	GDTRYTVSHNATQGNLNSLVGNAQATFNQATLNGNXS3GNASFNLSSNNAAQNGSLTLD	780	790	800	810	820	830
5	orf1.pep	NAKANVSHSALGNVNSLADKAIVFHPESSRETQISGGSKDTALHLKDSEWTLSGXELGNL	550	560	570	580	590	600
	orf1a	NAKANVSHSALGNVNSLADKAIVFHPESSRETQISGGSKDTALHLKDSEWTLSGXELGNL	840	850	860	870	880	890
10	orf1.pep	NLDNATITLNSAYREDAAGAQTGSATDAPRRRSRRSRSLLXVTPPPTSVESEFNTLTVNG	610	620	630	640	650	660
	orf1a	NLDNATITLNSAYREDAAGAQTGSATDAPRRRSRRSRSLLXVTPPPTSVESEFNTLTVNG	900	910	920	930	940	950
15	orf1.pep	KLNGGQGTFRPMSELGYRSDDKLKLAESSESGETYTLAVNNTGNEPASLFEQLTVVEGKDNDKPL	670	680	690	700	710	720
	orf1a	KLNQXGTFRPMSELGYRSDDKLKLAESSESGETYTLAVNNTGNEPVSQQLTVVEGKDNDKPL	960	970	980	990	1000	1010
20	orf1.pep	SENLNFTLQNEHVDAAGAW-----	730	740	750			
	orf1a	SENLNFTLQNEHVDAAGAWRYQOLIRKGDFEQLRHNPKVQEQLSDKLKGKAEEKKQAEAKDNAQS	1020	1030	1040	1050	1060	1070
25	orf1.pep	-----						
	orf1a	LDALIAAGRDAAEKTESVAEPARXAGGENVGIMQAEEKKRVRQADKDSALAKQREAETR	1080	1090	1100	1110	1120	1130
30	orf1.pep	-----						
	orf1a	XTTAFPRARKARRDLQPQPQPQPQPQRPDLXSRYSRANSGLSEFSATLNSVFAVQDELDR	1140	1150	1160	1170	1180	1190
35	orf1.pep	-----					760	--LDR
	orf1a	VFAEDRDRNAWVTSXIRXTKHYRSQDFRAYRQQTDLRQIGMQKNLGSGRVGVILFCHNRTE	1200	1210	1220	1230	1240	
40	orf1.pep	-----						
	orf1a	VFAEDRDRNAWVTSXIRXTKHYRSQDFRAYRQQTDLRQIGMQKNLGSGRVGVILFCHNRTE	1260	1270	1280	1290	1300	1310
45	orf1.pep	-----						
	orf1a	TFDDGIGNSARLAHGAVFGQYIGIDRFYIGISAGAGFSSGSLSDIGIGKXKRRRVLHYGIQA	830	840	850	860	870	880
50	orf1a	XFDDGIGNSARLAHGAVFGQYIGIDRFYIGISTGAGFSSGXLSDIGIGKXKRRRVLHYGIQA	1260	1270	1280	1290	1300	1310
55	orf1.pep	RYRAGFGGGIEPFHIGATRYFVQKADYRVEVNIAATPGLAFNRYRAGIKADYSFKPAQHI	890	900	910	920	930	940
	orf1a	RYRAGFGGGIEPFHIGATRYFVQKADYRVEVNIAATPGLAFNRYRAGIKADYSFKPAQHX	1320	1330	1340	1350	1360	1370
60	orf1.pep	SITPYXSLSYTDAA3GKVTRVNTAVLAQDFGKTRSAEWGVNAEIKGFTLSSLHAAAAGKP	950	960	970	980	990	1000
	orf1a	SITPYXSLSYTDAA3GKVTRVNTAVLAQDFGKTRSAEWGVNAEIKGFTLSSLHAAAAGKP	1380	1390	1400	1410	1420	1430
65	orf1.pep	OLEAQHSAGIKLGYRWX	1010	1020				
	orf1a	OLEAQHSAGIKLGYRWX	1440	1450				
70	The complete length ORF1a nucleotide sequence <SEQ ID 651> is:							

1 ATGAAAAACAA CGGACAAGC GACAACCGGA ACACACCGCA AAGCCCCGAA
 51 AACGGCCGC ATCCGCTTCT CGCCCTGCTTA CTTAGCCATA TGCCCTGTGCT
 101 TCGCATCTT CCCCCAACG TGGGGGGAC ACACATTATG CGGCATCAC
 151 TACCAAATCT ATCCGGACTT TGCGGAAAAT AAAGGCCAGT TTGCACTGCG
 201 GGCGAAAGAT ATGGAGGTNT ACACAAAAA AGGGGAGTTG GTCCGCAAT
 251 CAATGACAAAG AGCCCGATG ATTGATTTTG CTGTTGTTGTC GCCTAACGGC
 301 GTGGCGCAT TGGTGGCGA TCATATATG GTGAGCTGG CACATACCGG
 351 CGCTTATACG AACGGTGTATT TGTTGCGGA AGGAAGNAAAT CCCGATCAGC
 401 ACCGTTTTC FTACCAAAATT GTGAAAGAG ATAATTATAA GCCTGACAAAT
 451 TCACACCTTC ACAACCGGA ATACATATG CGCGCTTTCG ATAATTATTTG
 501 CACAGATGCA GAACCTGTCG AAATGACAGG TGACATGAGG GGGAAATACCT
 551 ATTCCGATAAA AGAAAATAAT CGCGACGCTG TCCCGATCGG CTCAAGGACAC
 601 CACTATGGC GTATGATG TGACAAACAC GCGGATTATG CTCATCTCCGG
 651 CGCTGATGTA ATTGCGCGCA ATACACATAT CGACGGTTGG GGAATAATG
 701 GCTANTTAG TTGAGCGGC GATGTCGCCG ATGCCAACGA CTATGCCCT
 751 ATGCGGATG CAGGTGGCGC AGGGACACGC GTTTGCCTAA TGTTTATTIA
 801 TGCAAAACAA ACACATTAAT GGCTGCTCAA CGGAGTTTTA CAAACGGCT
 851 ACCCTTATTC CGCGAGGAA AACGGTTTCG AGCTGATAGC CAAAGATTTG
 901 TTCTACGATC ACATTTACAG AGGGCATVACA CATACCGCTT NTTTGAACC
 951 CGCGATGACG GCGATTTT CTTTACATC ACACAAACAA GTGACCGGTA
 1001 CGGCTAACAGA AACCAACGAA AAGGTTTCACT ATCCAAAGCT TAAAGTACAG
 1051 ACAGTCGACG TTGTTGAGCA ATCTTGGAT GAAACTGATA AAGAACCGAT
 1101 TTACGCGGC GGGGGTTTA ATCGATACCGA TCCAAAGGTTA AACAACGGTG
 1151 AAAACCTTC TTATATCGAT TGCGGAAAC GCAAACCTCA CTATCAAC
 1201 AACATCAACG AAGGGGGGGG CGGGTTGAT TTTGAAAGGTTG ATTTTACCGT
 1251 CTGGCTGAC AACACAAACG CGTGGCAAGG CGCGGGGGT CATATCAGTG
 1301 AAGACAGTAC CGTTACTGG AAAGTAAACG CGGTGCAAAAG CGACCCCTG
 1351 TCCAAATTCG GCAAAAGCAC GTGCTGACGGT CAACCGAAAG GGGAAACCA
 1401 AGGCTGATG AGGCTGGGG CGCTGACAGT CATTTTGGAT CAGCAGGGAG
 1451 AGGATTAAGG CAAACAAACAA GCTTTTACTG AAATGCGCT GNTCAGCGGC
 1501 AGGGGTAAGG TCGCAACAAAG TGCGGATTAAT CAGTTCACCC CGACAAACACT
 1551 CTATTTCGGC TTGCGGGG GACGTTTGGG TTAAACCGG CATTGCGCTT
 1601 CGTTCACCCAT TATICAACAT CGGATGAGA GGGCGATGAT TGNCNATCAT
 1651 AATGGCCCAA CRACATCAC CGTACATTG ACAGGGATG AAAGTATTAC
 1701 ACACCGAGT GTGAAAGATA TCATAGACT TAATTACAGC AAAGAAATTTG
 1751 CCTACACGGC TTGTTTGGC GAGAAAGATA CGACCAAAAC GAACGGGGG
 1801 CTCACCTTG TTACCGCC CGCGCAGAAA GACCGCACCC NGCTGCTTTC
 1851 CGGCAGGAAAC ATTTTAAACG CGACATCAC GCAAACAAAC GGCAAAACTGT
 1901 TTTCAGCGG CGACAGGCA CGCCAGGCT ACAATCATTT AGGAAGCGGG
 1951 TTGTCACAAAG TTGGAAGGTAT CCCACAAAGGA GAAATCGTGT GGACAAACGA
 2001 CTGATCNAC CGACCGTTA AAGCGAAAAT TTCCATATG CAGGGGGGG
 2051 AGGCCGTGAT TTCCGGCAAT GTTGGCAAAAC TGGAAGGGGA TTGNCATTG
 2101 AACATCATCG CGGACAGGTT TTGTTGTCG CAACCGCATC AAAGGCATAC
 2151 AATCTGACG CTGTCGACT GGACNGTCTC GACAATTTG GTGGAANAAA
 2201 NCATTCACCGA CGTAAATTA TGACTTCAAT TGACTAAGAC NGACNTNAGC
 2251 GGCANTGNA NCNTNNCCNA TNACGNTNNNTN TAAANCTCN CNGGGNCNTGC
 2301 NNCACTNAAC GCGCATTTA GTGCAAAATGG CGATACAGCT TATACAGTCA
 2351 GCGCAACAGC CACCCAAAAC GGCAACCTTGA GCCTCTGTTG CAAATGCCAA
 2401 GCAACATTAA ATCAACGGCA ATTTAAACGGC ACACNATCGG NTICGGCAA
 2451 TGTCTATTG ATCTACGCA ACACGGCCG ACACGGCCG ACACGGCCG AGTCGAGGC
 2501 TTTCGCGCAA CGCTAACGGC PACGTAAGCC ATTCGGACT CAACGGCAAT
 2551 GTCTCCCTAG CGCGATTAAGGC AGTATTCCAT TTGAAARACG GCGCTTTCAC
 2601 CGGCAACACTC AGCGGCAAGCA AGGANACAGC ATTCACACTTA AAGACAGCG
 2651 AATGGACGCTT CGCTGCAAGC AGCGGATTAAG CGAAATTAAA CTTGACAC
 2701 GCCACCATTA ATCAACGGCA CGCTACGGTAC CGGCGGGCG CGGTTGCGGC CGTTCCTAT
 2751 AACGGCGAGG GTGTCAGACA CGCGCGGGCG CGGTTGCGGC CGTTCCTAT
 2801 TATCCGTATAC CGGCCACACT TCCTGAGAGAT CGCGTTTCA CAGCGTGA
 2851 GTAACGGCC AATTCGACAGC TCAAGGAAAC TTCCGCTTAA TGTCGGA
 2901 CTTCGGCTAC CGAACGGACA ATTGAGACTG CGCGGAAAGT TCCGAGGNA
 2951 CCTACACCTG GCGCTAACG ATACCGCCG ACGAACCGCT AAGCTCTGAT
 3001 CAATTCGAGG TAGTGGAGG AAAAGACACAC AAACCGCTGT CGGAAACACT
 3051 TAATTTCACG CTGCAAAACG AACACGCTGC TGCCGGGGG TGCGCTGTAC
 3101 AACATCATCGG CGAACGGCC GAGTTCCGGC TGCTAATATCC GGTCAAAGAA
 3151 CAAGAGCTT CGGACAAACG CGGCAACGCA GAAGCCAAA ACACGGCGA
 3201 AAAAGACACG CGGCAACAGC TTGAGCGGCT GATTGCGGGG GGGGGGGATG
 3251 CGCGCAAAA CGAACGAGCA AGTGGAGACTG CGCGGAAAGT TCCGAGGNA
 3301 GAAATGTCG CGTATTGCA CGGCGGGGA GGAGAAAACG GGTGCAAGGC
 3351 GGATTAAGACG AGCGGNTTGC CGAACACGG CGAACGGGA ACCGGGGCGG
 3401 NTACACCGC CTTCGGCGC GCGCGGGCG CGCGCGGGG TTGCGCGCAA
 3451 CGCGACGCC ACCGACGCC TCAACCCCAA CGCAGGGGG ACCTGTATNAG
 3501 CGGTATGCGC AATAGCGGT TGAGTGAATT TTCCGCGACG CTCAACAGCG
 3551 TTTTGGCGGT ACAGGACGGA TTGGACCGCG TGTTTGGCGA AGACCGCCG

3601	AACGCNGTTT	GGACAAGCG	CATCCGGNAC	ACCAAACACT	ACCGTTGCGA
3651	AGATTTCCGC	GCGCTACCGCC	AACAAACCGA	CCTGGCGCCTA	ATGGTATGCA
3701	AGAAAACCT	CGGCAGCGG	CGGCGCGGA	TCCIGTTTC	GCACAAACGG
3751	ACCGAAAACA	NCTTCGACGA	CGGCCATCGGC	AACTCGGCAC	GGCTTGCCCCA
5	3801	CGGCCCGCTT	TTGGCGCAT	ACGGCATCGG	CGAGTTGCGA
	3851	GCACCGGCGC	GGGTTTTAGT	AGCGGCANTC	TMTAGACCGG
	3901	AAAATCCGC	CGCCCGTGGT	GCATTACCGG	ATTCAGGCA
	3951	CGGTTTCCGC	GGATTTCGGCA	TGCAACCGTA	CATCGGGCA
10	4001	TCGTCACAAA	ACCGGATTA	CGCTACGAA	ACGTCATAAT
	4051	GGCTCTTGGT	TCAACCGNTA	CCNCGGGGG	ATTAAGGCG
	4101	CAACACGGCG	CAACACATNT	CTTACACNC	TTATTNNAGC
	4151	CCGATGCCG	TTCCGGCAAA	GTCCGAACAC	GGGTCAATAC
15	4201	GCTCAGGATT	TGCGCAACAA	CCCGAGTGG	GAAGGGGG
	4251	AATCAAAGT	TTTCACGCTGT	CCTNTCCACGC	TGCGCGCGGC
	4301	AACTGGAGC	GCACACACGC	CGGGCCTCA	AAATTAGGCTA
					CCGCTGGTAA

This encodes a protein having amino acid sequence <SEQ ID 652>:

1	MKTTDKRTTE	THR KAPK TGR	IRFS PAYLAI	CLSPFG IL PQA	WAGHTY FGIN
5	51	YQQYRDFAE	KGK FAVG ADR	I EYVN KKGEL	VGKS MTKAPM
10	101	VAALVQDQYI	VSVAHNGGY	NVDPEAEGEXN	IDFS VVS RNRG
15	151	SHPYNGDXHN	PRL KLFV TDRA	EPVEMTSMDM	GNTY SDKEKY
20	201	HWYRYDDKKH	GDL SYGS AML	IIGNH TMQG W	DTRV HAN DYG
	251	MPIAGAAGDS	GSP FMF YDRT	NNKWL LNVG I	QTGP YPS GRE
	301	FYDLYIRGDY	HFTVX FEP RSN	GHS FTS SNNN	GTGT VET TNE
	351	TVR LD ESDIN	ETD KEPV YA	GGV NQY RERL	KVSNPKL KVQ
25	401	NIN QGAGGLY	FGFD GTVSE P	NNTW QGAGV	H1S E DST VTV
	451	SK1 GKG TLHV	QAK GENO GSI	SVD GTV LIL	YNG K LIL SN
	501	RGT VQVL NADN	OQN PFD KLY LG	FRG RGR DL N	YSE D GAMI XX
	551	NATT STV T	TGNE ST QPS	HL SLS FHR I	QN D E KUT T KNG R
30	601	LALVQYPAAE	DR TXL SS GT	NLN GH I QT QN	GK L FFS GRT
	651	WSKMEGIPFQ	EIV WND WIX	RTFKAEN FHI	PHAY NH LCG S
	701	SNHAQAVFVG	AP HOS HTIC T	RSD WT GL TNC	I ASL T KTD X
	751	GK VXL XXXXX	XXL XG KAX LX	GNL SANGD U	ITV SHN ATON
	801	APT QM QAT L	NGX XSGN QAS F	NLS NRN AAQ NO	GNL SLV GNA Q
35	851	VSLADKA FV	FENS RT QQL	NGL SEFS T A	NLS VHS ALN GN
	901	AT ITL NSAY	HDA AGAQ TG	SDS EWT LPL SG	TEL GNL NLD N
	951	VNG KLN XQG	FR FMS E LFG	RSL LS T VPT	SVE SRN T LT
	1001	QLT VEG KGN	KPL SEN LNT	SE GTY TLA VN	NT GNE PV SLD
40	1051	QELS UKLGKA	EAKK QAEK DN	AQ SLD ALIA AR	GR DAA EKT ES
	1101	ENVG IM QAE E	EKK RV QD VA	SLA QK RE AE	VAE PAR XAG G
	1151	PQP QPO PO	PORD LK SRS Y	TR PKTTA FPR	AR XAR RD LP Q
45	1201	NAW TSX IR	T KHY RS QDR	AY QOT DLR Q	LDR VFA ED RR
	1251	TEN XFD DG	IS NAR L HAG	I PQ YGIG RFD	I GIST QAG FS
	1301	KIRR VFL HYG	I QARY RA FG	GFG IE PYI GA	TRY FVQ KAD Y
	1351	GLAF NRY RAG	I KAD YSF KP A	QH XS IT PY X S	LS YD AAS G K
	1401	AQDF GK TR SA	EWGV NAE IKG	FT LS XH AAA A	KGP QLE A QHS AG I KLG YRN *

A transmembrane region is underlined.

ORF1-1 shows 86.3% identity over a 1462aa overlap with ORF1a:

50	orfla.pep	10	20	30	40	50	60
		MKTTDKRTTE	THR KAPK TGR	IRFS PAYLAI	CLSF G IL PQA	WAGHTY FGIN	YQYYRDFAE
	orfl-1	MKTTDKRTTE	THR KAPK TGR	IRFS PAYLAI	CLSF G IL PQA	WAGHTY FGIN	YQYYRDFAE
		10	20	30	40	50	60
55	orfla.pep	70	80	90	100	110	120
		KGK FAVG AKD	I EYVN KKGEL	VGKS MTKAPM	IDFS VVS RNRG	V AAL VGD QY	I VSV AHNGGY
	orfl-1	KGK FAVG AKD	I EYVN KKGEL	VGKS MTKAPM	IDFS VVS RNRG	V AAL VGD QY	I VSV AHNGGY
		70	80	90	100	110	120
60	orfla.pep	130	140	150	160	170	179
		NVD FGA EGX NP	DQH RF Y QIV	KV RNN Y KPD NS	-HP YNG DX KMP RL HKF V	T DAE P VEM TS	X
	orfl-1	NVD FGA EG R	N DQH RF Y QIV	KV RNN Y KAG T KGH P Y GGD Y HMP RL HKF V	T DAE P VEM TS	Y M	
		130	140	150	160	170	180

	orfla.pep	180	190	200	210	220	230
		RGNTYSDKKEPYERVRIGSGHHYWRDDDKHGDL--SYSGR---			MLIGGNTNHMQGWGNN		
5	orfl1-1		DGRKYIDQNNYFDRVRIGAGRQYMRSDDEDPNRESSYHIAASAYSILVGGNTFAQNGSS				
		190	200	210	220	230	240
	orfla.pep	240	250	260	270	280	290
		GKVXSLSGD-VRHANDYGPMPFIAGRAAGDSSGSPMFYDVKTNKLWNLNVGLQTGYPSGRNG					
10	orfl1-1		GTGVNLGESEKIKHS--PYGFLPLTGGSGFDGSSGSPMFYDVKQWLWNLNVGLQTGPNYPIGKNSG				
		250	260	270	280	290	
	orfla.pep	300	310	320	330	340	350
		FOLIRKDWEYDDYRQDGDTHTVXFEPGRNSHGFSTSYNNNGTGTVTETNEKVSNP-KLKVT					
15	orfl1-1		: : : : : : : : : : :				
		FOLVRKDWFYDEBIFAGDTHSVFYEPRQNKGYSFNDDNNGTKINAKHEHNSLPNRLKTRT					
		300	310	320	330	340	350
	orfla.pep	360	370	380	390	400	410
		VRLFDESINLETDEKEPVY-AAGGVNQYPRPLRNNGENLFSIDYNGKLILSNNNNQGAGGLY					
	orfl1-1		: : : : : : : : : : :				
		360	370	380	390	400	410
20	orfla.pep	420	430	440	450	460	470
		FEFGDFTVSPENNETWQGAGHVHISEDSTTVWKVNGVANDRLSKIGKGTLLHVQAKGENQGSI					
	orfl1-1		: : : : : : : : : : :				
		FOGQDFTVSPENNETWQGAGHVHISEDSTTVWKVNGVANDRLSKIGKGTLLHVQAKGENQGSI					
25	orfla.pep	420	430	440	450	460	470
		420	430	440	450	460	470
30	orfla.pep	480	490	500	510	520	530
		SVGGDTVILQDQADDKGKKQAFSEIGLXSGRTVQLNADNOFPNDPKLYFGFRGRGLLDING					
35	orfl1-1		: : : : : : : : : : :				
		SVGGDTVILQDQADDKGKKQAFSEIGLXSGRTVQLNADNOFPNDPKLYFGFRGRGLLDING					
		480	490	500	510	520	530
	orfla.pep	540	550	560	570	580	590
		HSLSFHRIQNTDEGAMIXXHNATTITVTGNESTITQPSGNRNLNRSKEIAYNGWFG					
40	orfl1-1		: : : : : : : : : : :				
		HSLSFHRIQNTDEGAMIVHNHQDEKESTVITGNGNDIAT-TGNN-NSLDSKEIAYNGWFG					
		540	550	560	570	580	590
	orfla.pep	600	610	620	630	640	650
		EKDTTKTNGRNLNVYQPAEERTXLSSGTTNLNGNITQTNGKLFPSGRPTPHAYNHLSGG					
45	orfl1-1		: : : : : : : : : : :				
		EKDTTKTNGRNLNVYQPAEERTXLSSGTTNLNGNITQTNGKLFPSGRPTPHAYNHLSGG					
		600	610	620	630	640	650
	orfla.pep	660	670	680	690	700	710
		WSKMEGIPGEGIEIWWDNDWIXRTFKAEHNFIQGGQAVISRNVAKEVGDXHLSNHQAQFVG					
	orfl1-1		: : : : : : : : : : :				
		WSQKEGIPGEGIEIWWDNDWIXRTFKAEHNFIQGGQAVISRNVAKEVGDXHLSNHQAQFVG					
50	orfla.pep	660	670	680	690	700	710
		660	670	680	690	700	710
	orfl1-1	720	730	740	750	760	770
		APHQSHTICTRS DWTGLTNCVEXXITDKVIA SLTKTDXSGKVLXXXXXLXGXKAXLX					
55	orfl1-1		: : : : : : : : : : :				
		APHQSHTICTRS DWTGLTNCVEXKTTIDDKVIA SLTKTDISGNVLDLADHHLNLTGATLN					
		720	730	740	750	760	770
	orfla.pep	780	790	800	810	820	830
		GNLSANGDTRYT VSHNATQNGNLSLVGNQA T F N Q A T L N G N X S X G N A F N L S N N A A Q N G					
60	orfl1-1		: : : : : : : : : : :				
		GNLSANGDTRYT VSHNATQNGNLSLVGNQA T F N Q A T L N G N X S X G N A F N L S N N A A Q N G					
		780	790	800	810	820	830
	orfla.pep	840	850	860	870	880	890
		SLTLTSNDANAKVNSHSLN GN VS LADKAVPHFENSRFTGOLSGSX K T A L H L K D S E W T L P S G					
70	orfl1-1		: : : : : : : : : : :				
		SLTLTSNDANAKVNSHSLN GN VS LADKAVPHFENSRFTGOLSGSX K T A L H L K D S E W T L P S G					
		840	850	860	870	880	890

			900	910	920	930	940
	orfla.pep	TELGNLNLDNATITLNSAYRHDAAGAQGTGXVSDTPRRRSRRS--LLSVPPTSVESRFN					
5	orfl-1	TELGNLNLDNATITLNSAYRHDAAGAQGTGSATDAPRRRSRRSRRSLLSVTPPTSVESRFN	900	910	920	930	940
		950	960	970	980	990	1000
	orfla.pep	TLTVNGKLNKQGTFTRFMSELFGYRSDFKLKLAESSSEGTYTILAVNNTGNEPVSLQLTVVEG					
10	orfl-1	TLTVNGKLNGQGTFTRFMSELFGYRSDFKLKLAESSSEGTYTILAVNNTGNEPASLEQLTVVEG	960	970	980	990	1000
		990	1000	1010	1020	1030	1040
	orfla.pep	1040	1050	1060	KDKNKLPLSENLFNTLQNHEVHDAGAWRYQLIRKDGEFRLHNPVKEQELSDFKLGKAEEAKQAE		
15	orfl-1	KDKNKLPLSENLFNTLQNHEVHDAGAWRYQLIRKDGEFRLHNPVKEQELSDFKLGKAEEAKQAE	1020	1030	1040	1050	1060
		1060	1070	1080	1090	1100	1110
	orfla.pep	KDNQAQLDALIAAAGRDAEAKTESVAEPARKAGGENVGIMQAEEEKKRVQADKDTALAKQR					
20	orfl-1	KDNQAQLDALIAAAGRDAEAKTESVAEPARAOAGGENVGIMQAEEEKKRVQADKDTALAKQR	1080	1090	1100	1110	1120
		1130	1140	1150	1160	1170	1180
	orfla.pep	EASTRPXTTAFFPRAXXARDLPQPQPQPQPQPQPQDRLXSRYANSGLSSEPSATLNSVFV					
25	orfl-1	EASTRPXTTAFFPRAXXARDLPQPQPQPQPQPQDRLXSRYANSGLSSEPSATLNSVFV	1140	1150	1160	1170	1180
		1180	1190	1200	1210	1220	1230
	orfla.pep	QDELDLRVFAEDRNNNAWTSXIRXTKHYRSQDFRAYRQQTDLRQIIGMCKNLGSGRVGILFS					
30	orfl-1	QDELDLRVFAEDRNNNAWTSXIRXTKHYRSQDFRAYRQQTDLRQIIGMCKNLGSGRVGILFS	1200	1210	1220	1230	1240
		1240	1250	1260	1270	1280	1290
	orfla.pep	HNRNTENXFDDGIGNSARLAHGAvgFQGYGIGRFDIGISTGAGFSSGXLSDGIGGGKIRRVL					
40	orfl-1	HNRNTENXFDDGIGNSARLAHGAvgFQGYGIGRFDIGISTGAGFSSGXLSDGIGGGKIRRVL	1260	1270	1280	1290	1300
		1300	1310	1320	1330	1340	1350
	orfla.pep	HYGIQARYRAGFGGGIEPIYGIASTRYFVQKADYRYENVNVIATPGLAFNRYRAGIKADYSF					
45	orfl-1	HYGIQARYRAGFGGGIEPIYGIASTRYFVQKADYRYENVNVIATPGLAFNRYRAGIKADYSF	1320	1330	1340	1350	1360
		1360	1370	1380	1390	1400	1410
	orfla.pep	KPAQHXSITPYXSLSYTDAASGKVTRVRNTAVLAQDFGKTRSAEWGVNAEIKGFTLSXHA					
50	orfl-1	KPAQHXSITPYXSLSYTDAASGKVTRVRNTAVLAQDFGKTRSAEWGVNAEIKGFTLSXHA	1380	1390	1400	1410	1420
		1420	1430	1440	1450		
	orfla.pep	AAAKGPQLEAQHSAGIKLGYRWX					
55	orfl-1	AAAKGPQLEAQHSAGIKLGYRWX	1440	1450			
60	<u>Homology with adhesion and penetration protein hap precursor of <i>H.influenzae</i> (accession number P45387)</u>						
	Amino acids 23-423 of ORF1 show 59% aa identity with hap protein in 450aa overlap:						
	orfl 23	FXAAAYLAICLSFGFLIPQWAAGHTYFGINYQQYRDAENKGKFAVGAKDIEVYNKKGELVG	82				
65	hap 6	F + L C+S GI QWAAGHTYFGI+YQQYRDFAAEKGKF VGAK-IIEVYNK+G+LVG					
	orfl 83	FRNLFLACVSLGJIASQWAAGHTYFGIDQYQYRDFAAEKGKFVGAKNKEQQLVQ 65					
	hap 66	KSMTKAPMIDFSVSVSRNGVAALVGQYIVSVAHNGGYNVDGAEGRN-PDQHRTFYQIV 124					
	orfl 83	SMTKAPMIDFSVSVSRNGVAALVGQYIVSVAHNGGYNVDGAEGRN-PDQHRTFYQIV 142					
	hap 66	TSMTKAPMIDFSVSVSRNGVAALVGQDQYIVSVAHNGGYNVDGAEGRN-PDQHRTFYQIV 124					

Homology with adhesion and penetration protein hap precursor of *H.influenzae* (accession number P45387)

Amino acids 23-423 of ORF1 show 59% aa identity with hap protein in 450aa overlap:

	orfl 23	FXAAAYLAICLSFGFLIPQWAAGHTYFGINYQQYRDAENKGKFAVGAKDIEVYNKKGELVG	82
65	hap 6	F + L C+S GI QWAAGHTYFGI+YQQYRDFAAEKGKF VGAK-IIEVYNK+G+LVG	
	orfl 83	FRNLFLACVSLGJIASQWAAGHTYFGIDQYQYRDFAAEKGKFVGAKNKEQQLVQ 65	
	hap 66	KSMTKAPMIDFSVSVSRNGVAALVGQYIVSVAHNGGYNVDGAEGRN-PDQHRTFYQIV 124	
	orfl 83	SMTKAPMIDFSVSVSRNGVAALVGQYIVSVAHNGGYNVDGAEGRN-PDQHRTFYQIV 142	
	hap 66	TSMTKAPMIDFSVSVSRNGVAALVGQDQYIVSVAHNGGYNVDGAEGRN-PDQHRTFYQIV 124	

	orf1	143	KRNNYKAGTKGHPYGGDYHMPRLHLHKVXTDAEPVEMTSYMDGRKYIDQNNYPDVRIGAGR 202
			KRNNY+ A + HPY GDYHMPRLHLK VT+AEPV MT+ MDG+ Y D+ NYP-RVRIG+GR
5	hap	125	KRNNYQAWERKHPYGDYHMPRLHLHKFVTEAEPVGMTNMNDGKVYADRENYPERVRIGSGR 184
	orf1	203	QYWRSDEDEPNRRESSYHIA----- 222
			QYWRS+D+DE N SSY++
	hap	185	QYWRTDKDEETNVHSSYYVSGAYRYLTAGNHTQSNGNGTVNLSGNVSPNHGPPLTG 244
10	orf1	223	----SGSPMFIFYDAQKQKWILINGVLQTNPQYIGKSNGFQLVRKDWFDEIFAGDTHSVP 277
			SGSPMFIFYDA+K++WLIN VLQTG+P+ G+ NGFQL+R+WFY+E+ A DT SVF
	hap	245	GSKGDGSQSPMFIFYDAKKKQWLNALVQTLQTHGFGRNGFQLIREEFWYNEVLAVIDTPSVF 304
15	orf1	278	----YPEPRQNGFSNDNGTCKIN-AKEHENSLNLPNLKTRTVQFLNVSLSSETAREPVYHA 334
			Y P NG YF +N+GTGK+ + + + + TV-LFN SL++TA+V A
	hap	305	QRYTIPPINGHYSFVSNNNDGTGKLTLTRPSKDGSKASEVGTVKLFNPNSLNQTAKEHV-KA 363
20	orf1	335	AGGVNSYRPLRNNGENISFIDEKGGLRLLSTNSINQGAGGLYFGQDFTV-SPENNETWQGA 393
			A G N Y+PR+ G+NI D+GKG L + *INQGAGGLY+F+G+ F V +NN TWQGA
	hap	364	AAGHNYIYQFRMMEYKNYLGDQGKGTLLTENNINQGAGGLYFEGNFVVKGKQNNNTIWQGA 423
	orf1	394	GVHISEDSTVTWKVNGVANDRLSKIGKGT 423
			GV I +D+TV WKV+ NDRLSKIG GTL
	hap	424	GVSIGQDATVEWKVHNPEENDRLSKIGIGL 453
25	Amino acids 715-1011 of ORF1 show 50% aa identity with hap protein in 258aa overlap:		
	orf1	41	DTRYTVSHNATO-NGNXSLVXNAQATFNO-ATLNGNTSASGNASFNLSDHAVQNGSLTLS 98
			DT+ S TQ NG+ +L NA + A LNGN + + + F LS++A Q G++ LS
	hap	733	DTKVNISIPITQNGSINLTNNNATVNIHGLKLNQNVTLIDHSQFTLSNNNATQTGNIKLS 792
30	orf1	99	GNAKANVNSHSLALNNGNVSLLADKAVFHPESSRRTQQISGGKDTALHHLKDSEWLTPSGXELGN 158
			+A V+++ LNGNV L D R F ++S F QI G KDT + L+++ WT+PS L N
	hap	793	NHANATVNNTLNQVNLHLDQAQSFQSLKNSHFWHQIQGDQDTTVTLENATWTMFSDDTILQN 852
35	orf1	159	LNLNDNATITLNQSYRHDAGAQGATDAPXXXXXXXXXXXLLKVTPTSVESRNFNTLTVN 218
			L L+N-T+T+LNSAY + S+ +AP L PTS E RFNTLTVN
	hap	853	LTLNNSTVTLNNSAY-----SASSNNNPAHRHRS-----LETETTPTSAHRFNTLTVN 899
40	orf1	219	GKLNLQGQTFMSLPELGPRYSDKLKLQESSEGTYTAVLNNTGNEPASLQLTVVEGKDNKP 278
			GKL+GQTF+P S LGY+SDKLKL+ +EG YTL+V NTG EP +LEQLT+E+ DNKP
	hap	900	GKLNLQGQTFQTTSSLPGYQSKDLKLQSLNDAEGDYTTLVSRNTGKEPTLQLTLESLDNKP 959
	orf1	279	LSENLNFTLQNEHVDAGA 296
			LS+ L FTL+N+HVDAGA
	hap	960	LSDLKLKEPTLENDHVDAGA 977
45	Amino acids 1192-1450 of ORF1 show 41% aa identity with hap protein in 259aa overlap:		
	orf1	1	LDRVFAEDRRNAWTSGIRDTKHYRSQDFRAYRQQTDLRQIIGMOPNLGSGRVGILFSHNR 60
			LDR+ + + +AWTF+ +D + Y S FRAY+Q+T+L+RQIG+QK L +GR+G +FSH+R
	hap	1135	LDRFVDFQRAQSWNTNAQDKRYYDSAFRAYERQOKTNLQRQIGVQRALANGRIGAVFSHNR 1194
50	orf1	61	TENTFDGIGNSARLAHGAvgQGQYGDIFRYXXXXXXXXXXXXXIXGXKRRRVLYHG 120
			++NTFD+ + N A L + F QY K R+ +YG
	hap	1195	SDNTFDEQVNRHATLTMMSGFQAOYQWGLDQFGVNWGTGJISASKMAEEQSRKIHRAINYG 1254
55	orf1	121	IQARYRACFGGGFIIPIHAGATRYFVQKADRYENVNIAATPLGAFNFRYRAGIADYSFKPA 180
			+ A Y+ G GI+P+ G RYF+++ +Y+ E V + TP LAENRY AGI+ DY+P F P
	hap	1255	VNASYQFRLQGLQIOPYFGVNRYFIERENYQSEEVRKTPSLAENRYNAGIRVDYTPPT 1314
60	orf1	181	QHISITPYLSSLSTDAASGKVRTRVNTAVLAQDFGKTRSAEWNVAEIKGETLSLHAAAAA 240
			+IS+ Y+ P +Y+ D ++ V+T VN VL O PG+ E G+ AEI F +S + +
	hap	1315	1315 ISVSKPYPFFVNVYDVSNANQVQTIVNLTLVQZQPPGRYWQKEMVGLKAETLHFQISAFISKS 1374
	orf1	241	KGPOLAEQHSAGIKLGYRW 259
			+G QL Q + G+KLGYRW
65	hap	1375	QGSQSLGKQQNVGVKLGYRW 1393

Homology with a predicted ORF from *N.gonorrhoeae*

The blocks of ORF1 show 83.5%, 88.3%, and 97.7% identities in 467, 298, and 259 aa overlap, respectively with a predicted ORF (ORFing) from *N.gonorrhoeae*:

5	orf1.pep	MKTTDKRTTETHRKAPKTGIRIFXAAYLAICLISFGILPQAWAGHTYFGINYQQYRDFAE	60
	orf1ng	MKTTDKRTTETHRKAPKTGIRIFSPAYLAICLISFGILPQARAGHTYFGINYQQYRDFAE	60
10	orf1.pep	KGKFVAGAKDIEVYNNKGELVGKSNTKAPMIDFSVSRNGVAALVGQYIVSVAHNGYN	120
	orf1ng	KGKFVAGAKDIEVYNNKGELVGKSNTKAPMIDFSVSRNGVAALAGDQYIVSVAHNGYN	120
15	orf1.pep	NVDPGAEGXNIXDQXRXYTKIVKRNNYKAGTKGHPYGGDYHMPRLHXXVTDAAEPVEMTSY	180
	orf1ng	NVDPGAEGSN-PDQHRSFYQIVKRNNYKAGTNGHPYGGDYHMPRLHFKVTDAEPVEMTSY	179
20	orf1.pep	MDGRKYIDQNNYPPDRVRIGAGRQYWRSDEDPNNRESSYYIAS-----	223
	orf1ng	MDGWKYADLNKYFDVRIGAGRQYWRSDEDPNNRESSYYIASAYSWLVGGNTFAQNQSG	239
25	orf1.pep	-----GSPMFYIDAQQKQWLINGVLQZGNPYIYGKNSG	255
	orf1ng	GGTVNLGSEKIKHSYGFLPTGGSFGDGSGPMMFYIDAQQKQWLINGVLQZGNPYIYGKNSG	289
30	orf1.pep	FOLVRKDWFYDEIIFAGDTHSVFYPEPRQNGKYSFNDDNNTGKINAKHEHNSLPNRLLKTRT	315
	orf1ng	FOLVRKDWFYDEIIFAGDTHSVFYPEPRQNGKYSFNDDNNTGKIDAKHKHYSLPYRLLKTRT	359
35	orf1.pep	VQLFNVSLSETAREPVYHAAGGVNVSYRPLRNNGENISFIDEKGKELILTSNINQGAGGLY	375
	orf1ng	VQLFNVSLSETAREPVYHAAGGVNVSYRPLRNNGENISFIDEKGKELILTSNINQGAGGLY	
40	orf1.pep	FQGDPFTVSPEENNWTQAGWHISEDSTVTKVNGVANDRLSKIGKGT	422
	orf1ng	FEGNFTVSPRQNETWQGAGVHISDGSTVTKVNGVANDRLSKIGKGTLLVQAGKENQGSV	479
45	orf1.pep	DKVTASLTKTFDISGNVVDLADHAHLNLNTGLA	744
	orf1ng	FGVAPHQSHTICTRSWTGLTSTCERTITDDKVIASLSKIGKGTLLVQAGKENQGSV	774
50	orf1.pep	TLNQNLISANGCTR-YTSHNATQNGNXSLVXNAQATFTQATLNQNTSASGNASFNLSLDHA	803
	orf1ng	TFNGNL-VQAEATRTRLRRANATQNGNLISLVGNAQATFTQATLNQNTSASDNASFNLSDDNA	833
55	orf1.pep	VQNGSLTLSCNANAKANVSHSALGNVSLADKAFFHFESSRFTCQISGGKDOTALHLKDSEWT	863
	orf1ng	VQNGSLTLSDNANAKANVSHSALGNVSLADKAFFHFESSRFTCQISGGKDOTALHLKDSEWT	893
60	orf1.pep	LPSGKELGNNLNLDNATITLNSAYRHDAAGAQTSATDAPPRESRRSRSLLXVTPPTSVE	923
	orf1ng	LPSGTELGNNLNLDNATITLNSAYRHDAAGAQTSAADPAPPRESRRSLLSVPPTSVE	950
65	orf1.pep	SRENFTLTVNGKLNGQGTFRMSELEFGYRSRDKLKLAESESGTYTLAVNNTGNEPASLQLT	983
	orf1ng	SRENFTLTVNGKLNGQGTFRMSELEFGYRSRDKLKLAESESGTYTLAVNNTGNEPVSLEQLT	1010
	orf1.pep	VVEGKDNPPLSENLNFTLQNEHVDAQAW	1011
	orf1ng	VVEGKDNPPLSENLNFTLQNEHVDAQAWRYQLIRKDGEFRHLHNPKVEQELSDKLGKAGET	1070
	orf1.pep	-----	
	orf1ng	LDRVFAEDRRNAVWTSGIRDTHKYSRQDFR	1211
	orf1ng	PQRDLISKRYANSCLSEFSATLNSVFPQVAFDELDRVFAEDRRNAVWTSGIRDTHKYSRQDFR	1239
	orf1.pep	AYRQQTDLRQIGMQNLGSGRVGILFISHNRTENTFDDGIGNSARLAHGAVFGQYGRDFR	1271
	orf1ng	AYRQQTDLRQIGMQNLGSGRVGILFISHNRTGNTFDDGIGNSARLAHGAVFGQYGRDFR	1299

	orf1.pep	IGISAGACFSSCSLSDCICKXKRRRLVHYGIQARYRACFGGGFIEPHIGATRYFVQKADY 1331
	orf1ng	ICISACAGFSSCSLSDCIRGKIRRLVHYCQARYRACFGGGFIEPHIGATRYFVQKADY 1359
5	orf1.pep	RYENVNIAITPCLAFNRYRAGIKADYSFKAQHISITPYLSLSYTDAASGKVTRVNTAVL 1391
	orf1ng	RYENVNIAITPCLAFNRYRACIKADYSFKAQHISITPYLSLSYTDAASGKVTRVNTAVL 1419
10	orf1.pep	AQDFGKTRSAEWCVNAIHKFTLSSLHAAAKCPQLEQHQSAGIKLGYRV 1440
	orf1ng	AQDFGKTRSAEWGVNAIHKFTLSSLHAAAKCPQLEQHQSAGIKLGYRV 1468

The complete length ORF1ng nucleotide sequence was identified <SEQ ID 653>:

	1	ATGAAACCAA CCGACAAACG GACAACCGGA ACACACCGCA AAGCCCTAA
	51	AACCGCCGC ATCCGCTCT CGCCCGCTTA CTTAGGCCATA TGCCGTGCGT
15	101	TCGGCATCTT GCCCCAACGC CGGGCGGAC ACACATTATT CGGCATCAAC
	151	TACCAACTACT ATCCGCACTT TGGCCAAAAT AAAAGGCAAGT TTGAGCTGG
	201	GGGGAAGAT ATTGGAGTTT ACACAAAAG AGGGGAGTTG GTCCGAAAT
	251	CGATGCGAA AGCCCGCAAG ATTGAATTTC CTGTTGGTATC GCGTACCGGC
	301	GTGGCGCAT TGGCCGGCGA TCATATATT GTGAGCTGG CACATACGG
20	351	CGCGTATAC ATATGGTT TTGGTCGGG GGGAAAGCAAT CCCGATCACG
	401	ACCGCTTTC TTACCAAAAT GTAAAGAGAA ATAATTTATA AGCAGGGACT
	451	AACGGCCATC TTATGGCGG CGTATTCATC ATGCCGGTT TGACACAAATT
	501	TGTCACAGAT CGACAGACCTT ITGAGATGAC CAGTATATAGT GATGGTGGGA
25	551	ATAACCGCTGA TTAAATTAAC TACCTGTAC TTGTTGCAAT CGGAGCAGGC
	601	AGACARATTAT GGGCGCTGTA TGAAGACGAA CCCAAATAACCG GAAAGGTT
	651	ATATCATTTA GCAACGGCTT ATCTTGCGT CTCCGGTGGC ATACATTTTG
	701	CACAAATAGGC ATTCAGGGT GTGACAGCTCA ATCTGGTCTT CGGAAAGGATT
	751	AAACATAGCC CATATGGTT TTTACCAACG GGAGGCTCAT TTGGCGACAG
	801	TGGCTACCAAT GTTGTTCATCT ATGATGCCA ARAGCAAAAG TGTTTAATTAA
30	851	ATGGGGTATTG GCAAAACGGC AACCCCTATA TAGGAAAGG CAATGGCTTC
	901	CAGCTAGTC TTCAAGGNTG GTTCTATGAT GAAATCTTGG CTGGGAGATAC
	951	CCATTCTAGTA TCTACGAAAC CACATCAAA TGGGAAATAC TTTTTAACG
	1001	ACATATTAAC TGGGGCGG CAAATCTGTTT CAAACATTCAC ACACATATCT
	1051	CTACATATAA GTTAAAUAAAAC AGCAACGGTTT CAAATGTTTA ATGTTCTT
35	1101	ATCCGAGACA GCAAGAGACA CTGTTTATCA TGCTGCGGT GGGGGTCAACA
	1151	GTTATGACCA CAGACTGAAT ATGGAGAAA ATATTTCTT TATTGACAAA
	1201	GGAAAGGTTG AATTGATGAT TACACGACAC ATCACCAAGG GCGCGGGGG
	1251	TTTGTATTTT GAGGGTAAATT TACCGCTCTC GCCTAAAAAAC ARGAAAGCT
40	1301	GGCAAGGGC GGGCGCTTCA ATCAGTGTAC CGAGTACCGT TACTTGGAAA
	1351	GTAACAGGGC TGGCAACAGA CGCCGTGTC AAAATCGGCA AAGGCAACGCT
	1401	GCTGGTCTAA GCGAAAGGCG AAAAAACCGG CTGGGTGAGC GTGGGGGAGC
	1451	GTAAGCTCAT CTAGATCAG CAGGGCGGAT ATCAAGGCAA AAAACAAAGCC
45	1501	TTTAGTGAAT TGGCTTGGT CAGCGCAGG GGGACGGTGC ACTGAATGTC
	1551	CGATTAATCG TTCAACCCC CGAACACTCTA TTTCGGCTTT CGGGGGGAGC
	1601	TTGGGGATTG GAACGGCCAT TCGCTTCTGG TCCACCGAT TCAAAATACC
	1651	GATGGGGGG CGATGATGT CCACCCACAT CRAGACAAAG ATCCTAACCGT
	1701	TACCATACAA GCGAAATTAAG ATTAACTAC AACCCGCAAT AACAAACAACT
	1751	TGGTATGCAA AAAAGAAATTG GCTTACAAAGC TTGTTGTTGG CGAGAAAGAT
50	1801	CGACCAAAAG CGAACGGCC CTCAATCTG ATTACCAAC CGGAGAACG
	1851	GGATGCACT TTACTGCTT CCGGGCGAC AAATTTAACG GCGAATATCA
	1901	CGCAACAAAC CGCGAACACTG TTTTCAGCG CGACAGCCGAC ACCGCACGCC
	1951	TACATCATTT TAGGAAGCTG GTGGTCAAAAT ATGGAAGGTT TCCCCACAAAGG
	2001	AGAAATCTGTG TGGCAACAGC ATTGGATGCG CGGCACATT AAAGCGGAAA
55	2051	ACTTTCAAT TACGGGGGA AAACGGCTGG TTCCCGCA TAATGCGAAA
	2101	GTGGAAAGGGC ATTGGCATTT AAGCAATCAC GCCCCAAGCGAG TTTTCGGTGT
	2151	CGCACCGCAT CRAAGCCACA CAATCTGTAC ACCTTCGGC TGACCGGGTC
	2201	TGCAAGCTGTG TACCGGAAA ACATTACCG ACGATAAAAGT GATGGCTCA
	2251	TTGAGCAAGG CCGACATCG AGGCACATGCA ACCCTTGGC ATCACGCTCA
60	2301	TTAAATCTC ACAGGACTTG ACACACTCA CGGCAATCTT AGTCAGGGCG
	2351	GAGACACGCC CTATACGGT ACGGCAACGG CAACCCAAA CGCACCTTC
	2401	AGCTCTGGT GCAATGCCA AGCAACATT ATACRAAGCCA CATTAAACGG
	2451	CAACACATCG GCTTCGGACA ATCTCTATC TAATCTAACG AACACACGGCC
	2501	TACAAACGGC CGTCGTGACG TTTCGGCAAC CGCTTAAGGC AACAGTAAGC
	2551	CATTCCGGC TCAACGGCC TGTCTCTTA CGGCATTAAGG CGTATTCTCA
65	2601	TTTGAAGAAC AGCGCTTCA CGGGAAAGGAT CGCGGGGGC AAGGATACGG
	2651	CATTCAACTT AARAGRCAGC GAATGGCGC TGCCCTCGGG CACGGARTA
	2701	GGCAATTAAAC CCTTGACAA CGGCCACATT ACACCTCAATT CGCCCTATCG
	2751	ACAGCTGTGG CGAGGGCGGC AAACCGGAG TCGCGGAGAT CGCCCGCGCC
	2801	CGCGTTCGG CGGTCTTCA TTATCGTTA CGCGCCCAAC TTCCGGAGAA
70	2851	TCCCGTTCA ACACGGCTGAC GGTAACAGGC AAATGGACG GTCAGGGAAC

-371-

2901 ATTCGGCTT ATGTCGGAAC TCTTCGGCTA CGCGAGCGGC AAATTTGAAGC
 2951 TGGGAAAG TTCCGAAGGC ACTTACACT TGCGCTGCAA CAATACCGGC
 3001 AACGAAACCCG TAATGCTGA GCAATTGAGC GTAGTGGAAAG GAAAAGACAA
 3051 CACACCGCTG TCCGAATACT TTAACTTCAC CCTGCaaaAc gaacacgtcg
 5 3101 atgcggcgcg atggCGTTAT CAGCTTATCC qaaaaagacg CGAGTTCGc
 3151 CTGUAATACG CGGTCAAAAGA ACAGAGCTT TCCGACAAAC TCAGCAAGc
 3201 gggagaataACA GAggccgcgt TGAGGCCAA ACAGGCCaaC CTTCGCGCCA
 3251 AAcacaggc ggaaaaAGACG AACGcccaaa gcttgcggc gctgtatggc
 10 3301 gCggggcgc atgcacccga AAAGGCAGaa agtgttgcgc aacccGCCG
 3351 GCAGGGAGGC GGGGAAATcg ccgCATTTAT GCAGGGCGGAG GAAGAGAAAA
 3401 AACGGTGCAG GCGGATAAA GACACGCCCT TGCGCAAAAC CGCGGAAGGG
 3451 GAAACCCGGC CGCTGACCC TCTGGCTCCC CGCGCCGGCC CGGCCCGGG
 3501 GGATTPGCCG AACCCGAGC CCCAACCGCA ACCCCAAACG CAGCGGAC
 3551 TGATGAGCTT TTATGCCAAT AGCGTTTGA GTGAAATTTC CGGAGCGCTC
 15 3601 AACAGGGTT CGCCGCTACA GGACGAATTG GACCGCGTGT TTGCGGAAGA
 3651 CCGCCGCACG CGCTTGTGGA CAAGCGGCAT CGGGACGACC AAACACTACCC
 3701 GTTGCAGAAGA TTTCGGCGGC TACCGCCACAA AAACCGACCT CGCGCAAAATC
 3751 GGTATGCAAG AAAACCTCGG CAGCGGGCGC GTGCGCATCC TTGTTTCGCA
 3801 CAACCGGAGC GGAACACCTG TGACAGACG CATCGCAAC CGCGCACG
 20 3851 TTGGCCACGG TGCGCTTTC GGCGAAATAG CGATCGGCAG GTTGACATC
 3901 GGCGATCAGG CGGGCGCGG TTATGAGC GGCAGCGCTT CGAGCGCAT
 3951 CAGGGCAAA ATTCGGCCGC GGTGTCGCA TTACCGCATT CAGCGCAAGAT
 4001 ACCGGCGAGG CTTCGGCGGA TTTCGGCATCG AACCCACAT CGCGCAACG
 25 4051 CGCTATTTGC TCCAAAAGC GGATTACCGA TACGAAAACG TCAATATCGC
 4101 ACCGGCGGC CTTCGGATCA ACAGCTACCG CGCGGGCATT AAGGCAGATT
 4151 ATTCAATCAA ACAGCGGCAC CAATTTCA TCAACSCCTTA TTGAGCCTG
 4201 TCCATATTACCG ATGGCGCTTC CGGGAAAGTG CGAACGGGGC TCAATACCGC
 4251 CGTATTGGCG CAGGATTTCG CGAAAACCCG CGATGCGGAA TTGGGGCTAA
 4301 ACGGCGAAAT CAAAGGTTTC ACGCTGTCTCC TCCAGCGTGC CGCGGCCAAG
 30 4351 GGCGCCCAAT TGGAAGCGCA GCACAGCGCG GGCATCAAT TAGGCTACCG
 4401 CTGGTAA

This is predicted to encode a protein having amino acid sequence <SEQ ID 654>:

1 MKTTDKRTTE THRKAPEKTR IRFSPAYLAI CLSFGLPQKA RAGHTYFGIN
 51 YOYVDFRDAE KGFKEVAGKD IEVYNKKGEL VKGSKMTKAR IDFSVVSRRNG
 35 101 VAAALAGDDOI VSVAHNGGYN NVDFGAEGSN PDQHRSYQI VKRNNYKAGT
 151 NGHFGYGDYH MPRLHKFTD AEPMETYSYM DGWYKADLNK YPDVRVIRAG
 201 ROYWRSDDEDE PNNSREYSHI ASASWLYVG NTFAQNGSSG GTVNLSEKII
 251 KHRSYGFLLT GGSDFGSGSPE MFYDQAKOK WLJLNGVLYQG NEVYIGKSNF
 301 OVLRKDWFYD EITFGDHTHVF FYEHPHNGKY FFNDNNNGAG KIDAKHHKHS
 351 LPYRKLTRTV QLFNVNLSLET AREPVHYAAG GVNSYRPRLN INENISIFDK
 401 KGKELLTLST INQAGGGLYE EGNFTVSPKN NETWGAGVH ISDGSTVTWK
 451 VNGVANDRLS KIGKTGTLVQ AKGENGNGS3 VSGDGKVILQD QADDQGKKQA
 501 FSEIQLGVLSGR GTVQLNDAQNF FNFDKLYFGF RGGRDLINHG SLSFHRIQNT
 551 DEGAMVNHRQ QDKESTVTIT GNKDITTTGNN NNNLDSKEEL ANGNGWFGEKD
 601 ATTKNGLNLN NYPPFEADRT LLLSGGTNIA GNITCTNGKL FFSGRPTPH
 651 YNHLDGWSWK MEGIPOGEIV WDNWDWIDTR KAENFHIIQGG QAVVSRNRVAK
 701 VEGDWHLSNH AQAVGVEWQ OSHTICTRSD WTGLTSCTEK TTIDDKVIAS
 751 LSKTDVGRNVN SLADHAIHLN TGLATFNGNL VQAETrTIRL RANATONGNL
 801 SLVNGNQATF NQATLNGNTS ASDNAFNSL NNNAVONGSLT LSDNAKANVS
 851 HSALGNVNSL ADDKAVFHIEHN SFRFGKISGG KDTALHHLKDS EWTLPSTTEL
 901 GNLMNLNATI TLNSAYRHDA AGACTCSAAD APRRRLSRSRSL LSVTPPTSAE
 951 SRFTLTLTNG KLNQGQTFRF MSEFLGYRSQ KLKLAESSEG TTTLAVNNNTG
 1001 NEPVSLLEQPL VVEGKDNTPL SENLNFTLQN EHVDAGAWRY QLIRKOGEFR
 1051 LHNPKREQEL SDKLKGAKET EAALTAKQAO LAAQKQAEKD NAQSLDALIA
 1101 AGRNATEKAI SVAEPAQRAG NEAGIMQAE EEEKKRVQDAK DTALAKQREA
 1151 ETTRATTAF RARRARRDLP QPCQPQPOP QRDLISRYAN SGLSEFSATL
 1201 NSFVAVQDEL DRVFAEDDRN AWWTSGIRDY KHYRSQDFRA YRQOTDLRQI
 1251 GMQRNLGSQG VGILESHNRN GNTFDDGIGN SARIAHGAIVE QGYIGRFDI
 1301 GISAGAGFSS GSLSIDSGTRGK IRRRLVHGYI QARYRAFGGG FGIEPHIGAT
 1351 RYFVKQADYR YENVNIAITPG LAFNRYRAGI KADYSFKPAQ HISITPYLSL
 60 1401 SYTDAASGKV RTRVWLAQ DQFGKTRSAE WGVNAEIKGF TLSLHAAA
 1451 GPQLEAQHSA GIKLVGYRW*

Underlined and double-underlined sequences represent the active site of a serine protease (trypsin family) and an ATP/GTP-binding site motif A (P-loop).

65 ORF1-1 and ORF1ng show 93.7% identity in 1471 aa overlap:

		10	20	30	40	50	60	
	orf1-1.pep	MKTTDKRTTETHRKA PKTGRIRFSPAYLAICL SFGLPQWAQHHTYFGIN YQQYRDFEAEN						
5	orf1ng-1	MKTTDKRTTETHRKA PKTGRIRFSPAYLAICL SFGLPQARAGHTYFGIN YQQYRDFEAEN	10	20	30	40	50	
							60	
		70	80	90	100	110	120	
10	orf1-1.pep	KGKFAVGAKDIEV YNKKGELVGKSNT KAPNIDFSV SRNGVA ALVGDQ YIVSV VAHNGGYN						
orf1ng-1	KGKFAVGAKDIEV YNKKGELVGKSNT KAPNIDFSV SRNGVA ALAGDQ YIVSV VAHNGGYN	70	80	90	100	110	120	
		130	140	150	160	170	180	
15	orf1-1.pep	NVDPEGAER NPQDQHRT YKIVKRNN YKAGTK GHYPGG DYIMPR LKFV TDAP PEVENT SYN						
orf1ng-1	NVDPEGAER NPQDQHRT YQIVKRNN YKAGTK GHYPGG DYIMPR LKFV TDAP PEVENT SYN	130	140	150	160	170	180	
20	orf1-1.pep	DGRKYIDQNN YPDVR VIGAGR QWRS DEDP PNR RESS YHIA SAYS SWL VGG TAQ NGSG GG	190	200	210	220	230	240
orf1ng-1	DGWYKADLN KYPDVR VIGAGR QWRS DEDP PNR RESS YHIA SAYS SWL VGG TAQ NGSG GG	190	200	210	220	230	240	
25	orf1-1.pep	GTVN LGS SEK IKH PSY GFL PTGG SGFS GDS GPM FTY DAQ QCK WL ING VL QT GN PY IGK NS GF	250	260	270	280	290	300
orf1ng-1	GTVN LGS SEK IKH PSY GFL PTGG SGFS GDS GPM FTY DAQ QCK WL ING VL QT GN PY IGK NS GF	250	260	270	280	290	300	
30	orf1-1.pep	QLVR KDWF YEI FAG DTH VSY FEP PRQ NGK YSF NFD NN GK T KIN A HE NS LP N RL K TR TV	310	320	330	340	350	360
orf1ng-1	QLVR KDWF YEI FAG DTH VSY FEP PRQ NGK YSF NFD NN GK T KIN A HE NS LP N RL K TR TV	310	320	330	340	350	360	
35	orf1-1.pep	QLFN VSL SET ARE PVY HAAG GVSN YRPL LNNG ENIS FID DGK GEL LIL TSN INQ GAG LYF	370	380	390	400	410	420
orf1ng-1	QLFN VSL SET ARE PVY HAAG GVSN YRPL LNNG ENIS FID DGK GEL LIL TSN INQ GAG LYF	370	380	390	400	410	420	
40	orf1-1.pep	QGDFTV SPENN ETW WQAG HV H S E I G L V S G R G T V Q L N A D N Q F N P D K L Y F G R G R L D L N G H	430	440	450	460	470	480
orf1ng-1	EGNF TV SPENN ETW WQAG HV H S E I G L V S G R G T V Q L N A D N Q F N P D K L Y F G R G R L D L N G H	430	440	450	460	470	480	
45	orf1-1.pep	VGDGT VILD QQADD KGKK QAF SEI GLV S G R G T V Q L N A D N Q F N P D K L Y F G R G R L D L N G H	490	500	510	520	530	540
orf1ng-1	VGDGV VILD QQADD KGKK QAF SEI GLV S G R G T V Q L N A D N Q F N P D K L Y F G R G R L D L N G H	490	500	510	520	530	540	
50	orf1-1.pep	SLS FHRI QNT DEGA MIV VHN H N D K E S T V I T G N K D I A T T G N N L D S K K E I A Y N G F G E K D	550	560	570	580	590	600
orf1ng-1	SLS FHRI QNT DEGA MIV VHN H N D K E S T V I T G N K D I T T G N N L D S K K E I A Y N G F G E K D	550	560	570	580	590	600	
55	orf1-1.pep	TTKTN GRNL LVY QPA EE D T L L S G G T N L N G N T Q T N G K L F F S G R P T H A Y N H L D H W S Q	610	620	630	640	650	660
orf1ng-1	ATTKTN GRNL LVN YQPE EAD T L L S G G T N L N G N T Q T N G K L F F S G R P T H A Y N H L G W S K	610	620	630	640	650	660	
60	orf1-1.pep	KEGI PRGE IVW NDW D N I R T F K A E N F Q I K G Q A V S R N A V K G D W H L S N H A Q A V F G V A P H	670	680	690	700	710	720
orf1ng-1	MEGIP QGE IVW NDW D N I R T F K A E N F Q I K G Q A V S R N A V K G D W H L S N H A Q A V F G V A P H	670	680	690	700	710	720	
65	orf1-1.pep							
orf1ng-1								
70	orf1-1.pep							
orf1ng-1								

		730	740	750	760	770	780
5	orf1-1.pep	QSHTICTRSWTGLTNCVERTTIDDKVIASTLTKTDISGNVLADAHHLNLNTGLATLNGNL ::: : : : : : : : : : :					
	orf1ng-1	QSHTICTRSWTGLTSCTEKTITDDKVIASLSKTDIRGNVSLADAHHLNLNTGLATLNGNL 730 740 750 760 770 780					
10	orf1-1.pep	SANGDTTRYTVSHNATQNGNLSLVGNAQATFNQATLNGNTSAEGRNFNLSDHAVQNGSLT : : : : : : : : : : : :					
	orf1ng-1	SAGGDTHTYTVTRNATQNGNLSLVGNAQATFNQATLNGNTSAEGRNFNLSDHAVQNGSLT 790 800 810 820 830 840					
15	orf1-1.pep	LSGNAKANVSHSALGNVSLADKAVFHPESSRFTQKISGGKDTALHLKDSEWTLPSCTEL : : : : : : : : : : : :					
	orf1ng-1	LSDGNANVSHSALGNVSLADKAVFHPESSRFTQKISGGKDTALHLKDSEWTLPSCTEL 850 860 870 880 890 900					
20	orf1-1.pep	GNLNLDNATITLNSAYRHDAAGAQQTGSATDAPRRRSRRSRSLLSVTPTSVESRFNTLT : : : : : : : : : : : :					
	orf1ng-1	GNLNLDNATITLNSAYRHDAAGAQQTGSADAPRRRSR---RSSLSSVTPTSVESRFNTLT 910 920 930 940 950 960					
25	orf1-1.pep	VNGKLNGQGTFRFMSELPGYRSRDLKLAESSEGTYTLAVNNNTGNEPASLEQLTVEGKDN : : : : : : : : : : : :					
	orf1ng-1	VNGKLNGQGTFRFMSELPGYRSRDLKLAESSEGTYTLAVNNNTGNEPVSLEQLTVEGKDN 960 970 980 990 1000 1010					
30	orf1-1.pep	KPLSENLNFTLQNNEVHDAGAWRYQLIRKDGEFRLHNPKVEQELSDKLKGKA----- : : : : : : : : : : : :					
	orf1ng-1	TPLSENLNFTLQNNEVHDAGAWRYQLIRKDGEFRLHNPKVEQELSDKLKGAKETEAALTAK 1020 1030 1040 1050 1060 1070					
35	orf1-1.pep	----BAKKQAEKDNQAQSLDALIAAGRDAVEKTESVAEPARQAGGENVGIMQAEZEKKRVO : : : : : : : : : : : :					
	orf1ng-1	QAOAAKQQAEKDNQAQSLDALIAAGRGRATEKAESPAQAGGENAGINQAEZEKKRVO 1080 1090 1100 1110 1120 1130					
40	orf1-1.pep	1130 1140 1150 1160 1170 1180 ADKDATALAKQKREAEETRATTAFPRARRRARRDLQQLQQPOPOPOQQRDLISKYANGLSEFS : : : : : : : : : : : :					
	orf1ng-1	1140 1150 1160 1170 1180 1190 ADKDATALAKQKREAEETRATTAFPRARRRARRDLQQLQQPOPOPOQQRDLISKYANGLSEFS : : : : : : : : : : : :					
45	orf1-1.pep	1190 1200 1210 1220 1230 1240 ATLNSVFAVQBDLDRVFAEDRNRNAVWTSIGERDTKHYSQDFRAYRQQTDLRQIGMQKNLG : : : : : : : : : : : :					
	orf1ng-1	1200 1210 1220 1230 1240 1250 ATLNSVFAVQBDLDRVFAEDRNRNAVWTSIGERDTKHYSQDFRAYRQQTDLRQIGMQKNLG : : : : : : : : : : : :					
50	orf1-1.pep	1250 1260 1270 1280 1290 1300 SGRVGILFSHNRRTENTFDGIGNSARLAHGAVFGQYGIIDRFYIGISAGAGFSSGSLSDGI : : : : : : : : : : : :					
	orf1ng-1	1260 1270 1280 1290 1300 1310 SGRVGILFSHNRRTGNTFDGIGNSARLAHGAVFGQYGIIDRFYIGISAGAGFSSGSLSDGI : : : : : : : : : : : :					
55	orf1-1.pep	1310 1320 1330 1340 1350 1360 GGKIRRRVLHYGICQARYRAGFGGGFGEPIHIGATRYFVQKADYRKYENVNIAATPGLAFNRYR : : : : : : : : : : : :					
	orf1ng-1	1320 1330 1340 1350 1360 1370 RGKIRRRVLHYGICQARYRAGFGGGFGEPIHIGATRYFVQKADYRKYENVNIAATPGLAFNRYR : : : : : : : : : : : :					
60	orf1-1.pep	1370 1380 1390 1400 1410 1420 AGIKADYSFKPAQHISITPYLSLSYTDASGKVTRVNTAVLAQDFGKTRSAEWGVNAEI : : : : : : : : : : : :					
	orf1ng-1	1380 1390 1400 1410 1420 1430 AGIKADYSFKPAQHISITPYLSLSYTDASGKVTRVNTAVLAQDFGKTRSAEWGVNAEI : : : : : : : : : : : :					

-374-

	1430	1440	1450
orf1-1.pep	KGFTLSLHAAAAGKGPQEAQHSAGIKLGYRWX		
orf1ng-1	KGFTLSLHAAAAGKGPQEAQHSAGIKLGYRWX		
	1440	1450	1460

In addition, ORF1ng shows 55.7% identity with hap protein (P45387) over a 1455aa overlap:

		600	610	620	630	640	650	
	orflng-1.pep	FKDATKTNGLRNLYNQPEADRTLSSLGGTNLNGNITQTNGLKFFSGRPTPHAYNHLGSG						
	p45387	: :						
5		ETDKNKHNGRLNLITYKPTTEDRTLLISGGTNLKGDDITQTKGKLFSGRPTPHAYNHLNKR	570	580	590	600	610	620
		660	670	680	690	700	710	
	orflng-1.pep	WSKMEGIPQGEIVWDNDWIDRTFKAENFHICQQGQAVVSRNVAKVEGDWHLNSHAQAVFGV						
10	p45387	:						
		WSEMEGIPQGEIVWDHWINRTFKAENFIQKGGSAVVSRRNVSIEGNWTVSNNANATFGV	630	640	650	660	670	680
		720	730	740	750	760	770	
15	orflng-1.pep	APHSQSHTCTRSWDTGLTSCTEKTIDDKVIAISLSKTDIRGNVSLADAHHLNTGLATIN						
	p45387	:						
		VPNQNTCTCRSDWDTGLTTCQKVLDLTDKVIINSIPKTQINGSINLTNDNATANVKGALKN	690	700	710	720	730	740
		780	790	800	810	820	830	
20	orflng-1.pep	GNLSSAGGDHTYHTVTRNATQNQNLNSLVGNAQATFNQATLNQNTSASDNASFMNLSNNAVQNG						
	p45387	:						
		GNVTL-----TNHSQFTLSNNATQIG	750			760		770
		840	850	860	870	880	890	
25	orflng-1.pep	SLTLSDNAKANVSHSALNNGNVLSDLAKVAFHFFENSRTFGKISGGKDOTALHLKDSEWTLPSG						
	p45387	:						
		NIRLDNSTATVDVNANLNGNVLTSQAFSLKNSHFSHQIQGDKGTTVTLENATWTMPSD	780	790	800	810	820	830
		900	910	920	930	940	950	
30	orflng-1.pep	TELGNLNLDNATITLNSAYRHDAAGAQGTSAAADAPPRRRSRSSLSTVPPTSAESRFNLT						
	p45387	:						
		TTLQNLTLNNSTITLNSAY----SASSNNTPRRRS---LETETTPTSAEHRFNTLT	840		850	860	870	
		960	970	980	990	1000	1010	
40	orflng-1.pep	VNGKLNGGTFRPMSELFGYRSQGSKLKLLESSETCYTLAVNNTGNEPVSLQELTVEVGKD						
	p45387	:						
		VNGKLGSQGTFQFTSSLFGYKSDFKLKLNSNDAEGDYILSVVRTGKEPETLEQLTJVESKD	880	890	900	910	920	930
		1020	1030	1040	1050	1060	1070	
45	orflng-1.pep	TPFLSENINFTLQNEHVDAAGAWYQLKURDGSEFRHLNFPVKEQELSOKLGKAGETEAALTAK						
	p45387	:						
		QPLSDLKFLTLENDHVDAGLARYKLVKNDEGFRHLNPKEQELHNLDLVRAEQAERTLEAK	940	950	960	970	980	990
		1080	1090	1100	1110	1120	1130	
50	orflng-1.pep	QAGLAAKQQAEKQDQAQLDLAIAAGRNTAEKAESVAEPARQAGGENAGIMQAEEEKKRV						
	p45387	:						
		QVEPTAKTQTGEPKVRSRRAARAFFPDTLPDQSLLNAALEAKQAE-LTAETQSKSAKTKKV	1000	1010	1020	1030	1040	1050
		1140	1150	1160	1170	1180	1190	
55	orflng-1.pep	QAKDK--DTALAKQREAETRFATTAFPRARRRD-LFQFQPOPQPFQFQDRLSLRYANSG						
	p45387	:						
		RSKRAVFSDPLLSQSLFALEALEVIDAPCQSEKDRLACEEAEKQ-RKQKDLSLRSYNSA	1060	1070	1080	1090	1100	1110
		1200	1210	1220	1230	1240	1250	
60	orflng-1.pep	LSEFSATLNSVFAVQDDELDRVFAEDRNRNAVWTSGIRDTKHYRSQDFRAYRQQ-TDLRQIG						
	p45387	:						
		LSELATVNSMSLVSQDEDLRFVDDQAGSAWVNTNIAQDKRKYDSDAFRAYQQNTLNQIG	1120	1130	1140	1150	1160	1170
		1260	1270	1280	1290	1300	1310	
65	orflng-1.pep	MQKNLGSGRGVGILFSHNRNTGNTGFDIGNSARLANGAVFGQYQIGRFDIGISAGGFSSG						
	p45387	:						
		VQKALANGRIGAVFSHSRSDNTFDEQVKHATLTMMSGFAQYQWGLQFGVNVTGJISAS	1180	1190	1200	1210	1220	1230

-376-

		1320	1330	1340	1350	1360	1370
	orfing-1.pep	SLSDGIRKIGKIRRRLHYGICQARYRAGFCGGFIEPHIGATRYFVQKADYRYEVNNTATETGL					
5	p45387	: ::: : : : : : : : : : : : :	KMAEAEQSRSKIHRRKAINYGVNASYQFRLGQLGIQFYFGVNRYFIERENYQSEEVRVKTPSL				
		1240	1250	1260	1270	1280	1290
		1380	1390	1400	1410	1420	1430
10	orfing-1.pep	AFNRYYRAGIKADYSFRKPAQHISITPYLSSLTYDAAASGKVTRVNTAVLAQDFGKTRSAEW					
	p45387	: ::: : : : : : : : : : : :	AFNRYYNAGIRVDTFTPTDNISVKPYFFFNYDVDSNANVQTTVNLTVLQQPFGRYWKQEV				
		1300	1310	1320	1330	1340	1350
		1440	1450	1460	1469		
15	orfing-1.pep	GVNAEIKGPFTLSHAAAAKGPQLEAQHSAGIKLGYRWX					
	p45387	: ::: : : : : : : : : : :	GLKAELIHLFQISAFISKSQGSQLGKQNVGVKLGYRW				
		1360	1370	1380	1390		

Based on this analysis, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

20 Example 78

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 655>:

1	..AAGGTGTCGC	AATTTGTCGA	AGA,CCGCG	CGGNGCCGTCG	TGCCCTGCGCA
51	CAGTTTTGGA	CCGACCGCGC	AAAATTGGA	CCCTTTTAAG	GGGGGTGCGG
101	CAACCCATT	GTTTATGAA	GATCAAAATG	TCTGCAAAGG	TTTGCCAGGG
151	CAGTTCCTCG	CTTATGCCG	TAACTTCCCG	GTTTGGCCG	ATCAGGCAA
201	CGCGATGTG	CAGATGCGC	TTTGAGACG	ACTTGGCGG	GTGGCGTAG
251	GTGCAAACT	GCACATTCAC	AATCCCTTGC	CCGAATGCCG	GATTGCCAA
301	GGCTGGATAA	TCCCCAAAC	CTGGTTGTC	CGGGCACAA	TGGTTATCGG
351	CGGTTATGAA	GGGGGGCGAG	GTGAAAAGAC	CITTGAACC	GPTGCAGAAC
401	GTTCGAAGT	GTTCGGCGCA	TA		

This corresponds to the amino acid sequence <SEQ ID 656; ORF6>:

1	..KWWQFVEXPL	RAVVFPADSFE	PTAQLKNLF	AGAATILFY	DQNQVVKGLQE
51	QFPAYAAANF	VWADQDNAM	VQAYAWTTLA	VVGVGNLQHY	NPLPDAAIK
101	AWNIPENWL	RQJMIVIGIE	GAAGEKTFEP	VABRLKVFGA	*

35 Further sequence analysis revealed a further partial DNA sequence <SEQ ID 657>:

1	..CTCGCTGCC	CTCGCTCTG	CGACAGTTT	GAACCGACCG	CGCAAAATT
51	GAACCTGTT	AAAGCCGGTG	CGCGAACAT	TTTGTGTTAT	GAAGATCAA
101	ATGTCGTCRA	AGGTGGCAG	GAGCAGTTCC	CTGCTTATGC	CGCTAACTTC
151	CCCGTTGGG	CGGATCAGG	AAACCGCATG	GTGCACTATG	CGGTGCGAC
201	GACACTTGGC	GGCGTCGGCG	TAGGTGCAA	CTTGGCAACAT	TACATTCCT
251	TGCCGATGTC	GGGGATGCG	AAAGCGTGA	ATATCCCGA	AAACTGTTG
301	TTGGGGCGAC	AAATGGTTAT	CGGGGGTATT	GAAGGGGGGG	CAGGTGAAAA
351	GACCTTTGAA	CCGGTTGCGAG	AACCTTTGAA	AGTGTTCGGC	GCATAA

This corresponds to the amino acid sequence <SEQ ID 658; ORF6-1>:

45	1	..LRVQVLPDSE	PTAQLKNLF	KAGATILFY	EDQNVVKGLQ
	51	PVWAQDNAM	VQAYAWTTLA	VVGVGNLQH	YNPLPDAIAK
	101	AWNIPENWL	RQJMIVIGIE	GAAGEKTFEP	VABRLKVFG A*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

50 ORF6 shows 98.6% identity over a 140aa overlap with an ORF (ORF6a) from strain A of *N.meningitidis*:

-377-

	orf6.pep	KVWQFVEXPLRAVVVPADSFEPTAQKLNLFK
5	orf6a	Q1IVEHAVLHTPSSFNQSARVVVLFGEEHDKVWQFVEDALRAVVVPADSFEPTAQKLNLFK 40 50 60 70 80 90
	orf6.pep	AGAAATILFYEDQNQVVKGLQEQQFPAYAANFPVWADQANAMVQYAVWTTLAAVGVGANLQHY
10	orf6a	AGAAATILFYEDQNQVVKGLQEQQFPAYAANFPVWADQANAMVQYAVWTTLAAVGVGANLQHY 100 110 120 130 140 150
	orf6.pep	100 110 120 130 140 NPLPDAATAKAWNIPENWLRLRAQMVGIGIEGAAGEKTFEPVAERLKVFGAX
15	orf6a	NPLPDAATAKAWNIPENWLRLRAQMVGIGIEGAAGEKTFEPVAERLKVFGAX 160 170 180 190 200

The complete length ORF6a nucleotide sequence <SEQ ID 659> is:

1	ATGAGCCGTG AATCTCTGCA ACAGGGCTGC GAAAGCCGCC GTTCCATTAA
51	TTCCTTAAT AAAAACCTGC CGCTCGGCCA AGATGAAATC GTCCAATTCG
101	TGGAACACGC CGTTTTGAC ACACCTCTT CGTTCAATTC CCAATCTGCC
151	CGTGGTGTGC TGCTGTGTC CGAAAGAGCAT GATAAGGTGT GGCAATTTTG
201	CGAACAGCGG CTGCGTGCCTC TGCTGCTGC CGACAGTTT GAACGCCACCG
251	CGAACAAAATG GAACTCTGTTT AAAGCCGGTG CGGCAACTATAT TTGTTTTAT
301	GAAGATCAAATGCTGCTAA AGGTGTCAG GAGCAGTCC CTGCTTATGC
351	CGCCAACTTCC CGCGTTGCG CGGACAGCGG GAACCCGGATG GTGCGATGATG
401	CGGTTGTGAC GACACTTGC CGGGTCGGCG TAGGTGCAAC CTCGCAACAT
451	TACACATCCC TGCGCGATGCG CGCGATTTGCA AAAGCGTGGAA ATATCCCCGA
501	AAACTGGTTG TTGCGCGCAC AAATGGTTAT CGGGCGTATT GAAGGGCCGG
551	CAGGTAAAAA GACCTTTGAA CCAGTTGCG AACGTTGAA ACTGTTGCGC
601	GCATAA

This is predicted to encode a protein having amino acid sequence <SEQ ID 660>:

1	MTRQSLQQA ESRSLISLYIN KNLPVGKDEI VQ1IVEHAVLH TPSSFNQSA
51	RVVVIIFGEEH DKWVQFVEDA LRAVVVPADS EPTAQKLNLF KAGAAATILFY
101	EDQNQVKGLQ EQFPAYAANFPVWADQANAM VQYAVWTTLA AVGVGANLQHY
151	YNPLPDAIA KAWNIPENWL RLRAQMVGIGIEGAAGEKTFE PVAERLKVF
201	A*

ORF6a and ORF6-1 show 100.0% identity in 131 aa overlap:

40	orf6a.pep	50 60 70 80 90 100 TPSSFNQSA RVVVIIFGEEH DKWVQFVEDA LRAVVVPADS EPTAQKLNLF KAGAAATILFY orf6-1 LRAVVVPADS EPTAQKLNLF KAGAAATILFY 10 20 30
45	orf6a.pep	110 120 130 140 150 160 EDQNQVKGLQ EQFPAYAANFPVWADQANAM VQYAVWTTLA AVGVGANLQHY NPLPDAIA
	orf6-1	EDQNQVKGLQ EQFPAYAANFPVWADQANAM VQYAVWTTLA AVGVGANLQHY NPLPDAIA 40 50 60 70 80 90
50	orf6a.pep	170 180 190 200 KAWNIPENWL RLRAQMVGIGIEGAAGEKTFE PVAERLKVF orf6-1 KAWNIPENWL RLRAQMVGIGIEGAAGEKTFE PVAERLKVF 100 110 120 130

Homology with a predicted ORF from *N.gonorrhoeae*

ORF6 shows 95.7% identity over a 140aa overlap with a predicted ORF (ORF6ng) from *N.gonorrhoeae*:

-378-

	orf6.pep	KVWQFVEXPRLRAVVPADSFEPTAQKLNLFK	30
	orf6ng	SNVSLDMNSNPTVLRMGLPLVIASLRRGAIYKVWQFVEDALRAVVPADSFEPTAQKLKLFK	64
5	orf6.pep	AGAAATILFYEDQNVVKGLQEOPFPAYAANFPWADQANAMVQYAVWTTLAAVGVGANLQHY	90
	orf6ng	AGAAATILFYEDQNVVKGLQEOPFPAYAANFPWADQANAMVQYAVWTTLAAVGAGANLQHY	124
10	orf6.pep	NPLPDAIAKAWNIPENWLLRAQMVGIGGIEGAAGEKTFEPVAERLKVF	140
	orf6ng	NPLPDVAIAKAWNIPENWLLRAQMVGIGGIEGAAGEKVFEPVAERLKVF	174

The complete length ORF6ng nucleotide sequence <SEQ ID 661> was identified as:

15	1 ATGGCGGTG CGTCAAATGT CAGCTTGGAT CTACGGTGT 51 51 ACGCTGGGA TTACCTTAT ATATTGCGTC CCTTAAGAGG GCGCCCAATAT
	101 101 ATAAGGTGT GCAATTGTC GAAGACGC CG TGCGTGCCTG CTCGCTTGCGC
	151 151 GACAGTTTG AACCGACCG GC AAACAAATTC AAAGCTGTATA AGGCCGGCCG
	201 201 GGCACCAT TTTTTATTG AAGATCAAA TGTCGTCAAA GTTTCAGG 251 251 AGGACTTCCC TGCTTATGCC GGCACATTTC CGGTGTTGGC GUACAGGG 301 301 AACCTTATGC TCACTATGC CGCTGGAGG AACACTTGC CGCGCGGTGC
	351 351 AGGCTGAAAT CTGCAACATT ACAACCCCTT GCGCGATGTG CGGATTGTCA 401 401 AAGCGTGGAA TATTCGGAA AACITGGCTT GC CGGGCGCA AATGTTATC 451 451 GGTGTTATTG AAAGGGCGCG aggtggaaaa gtcttggaaac CCCTTGCgga 501 501 acgtttgAAA GTCTTGGCG CATAA

25 This encodes a protein having amino acid sequence <SEQ ID 662>:

15	1 MAVASNVSLSD MSNPVTLRMRG LPLYIASLRR GAIYKVNQFV EDALRAVVP 51 51 DSFEPTAQKL KLFKAGAATI LFYEDQNVVK GLQEOPFPAYA ANFPWADQ 101 101 NAMVQYAVWT TLAAVGAGAN LQHYNPLPDV AIAKAWNIPE NWLLRAQMVI 151 151 GGIEGAAGEK VFEPVAERLK VFCA*
----	---

30 ORF6ng and ORF6-1 show 96.9% identity in 131 aa overlap:

			10	20	30
	orf6-1.pep		LRAVVPADSFEPTAQKLNLFKAGAATILFY		
35	orf6ng	PTVLRMGLPLVIASLRRGAIYKVWQFVEDALRAVVPADSFEPTAQKLKLFKAGAATILFY			
		20 30 40 50 60 70 80 90			
40	orf6-1.pep	EDQNVVVKGLQEOPFPAYAANFPWADQANAMVQYAVWTTLAAVGVGANLQHYNPLPDAIA orf6ng	EDQNVVVKGLQEOPFPAYAANFPWADQANAMVQYAVWTTLAAVGAGANLQHYNPLPDAIA 80 90 100 110 120 130		
45	orf6-1.pep	KAWNIPENWLLRAQMVGIGGIEGAAGEKTFEPVAERLKVF orf6ng	KAWNIPENWLLRAQMVGIGGIEGAAGEKVFEPVAERLKVF 140 150 160 170	100 110 120 130	

50 It is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 79

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 663>

55	1 ..GGCTACAAT ACTCTTCCG CGCGGGCAGC CGCATGCCA ACTACCAAT 51 51 CACCGCGATC CCTCTTGC CG ACCGGCTGCG CGATACGGGT CAATGCCAAC 101 101 AC CGCGCGCT ATGAGCGCGCT AGAACTCGTG CGCGCGGTG CGGGCGCTG 151 151 GGACGGCAGG CGCGAGCTT CCCCACCGT CAATCTGGTG CGCAAACGCC 201 201 TGGCCCGCAA CCCTTGTCTT GAAGTCCCG CGGAAGCGGG CAACCGGAAA
----	--

-379-

251 CATTGCGGC TGGACCCGGA CGTATCGGGC AGCCTGAACA CGGAAG. ctc
 301 rCTGGCGgGC CGCCTGTTT CCATACTTCG ACCGGGCGAC TGCTGGGGC
 351 GGCGCARGC CAGCGCsAT CGCCGACTCT ACGGCATTTT GGAATACGAC
 401 ATCGCACCGC AAACCCGGT CGAACCAcGC ATGGACTACC AGCAGCGAA
 451 AGAAAACGCC GACGGCGGC TAGAGTAACG CTGTGACAC ACCCAAGGT
 501 ATGGCACCGC CTTCGGCCCG AAAGACACCC CGCCACAAA TTGGCGAAC
 551 AGCCACCAAC GTGCGCTAA CCTGTGCGG GGCATCGAC ACCGCTCAA
 601 CGAAGACTGG AAACCTAAAG CGGAATACGGA CTAC..

This corresponds to the amino acid sequence <SEQ ID 664; ORF23>:

10 1 . . . GVNLYFLFARG RIANYQINGI PVALADALTG NANTAAAYER EVVRGVAGLL
 51 DGTGEPSATV NLVRKLRLTR KLFEVRAEAG NRKHGFLDAD VSSGLNTEXX
 101 RTGRLVSTFG RGDSWRRER SRXAAELVGLI EYDIAPQTRV HAXMDYQQA
 151 ETADAPLSYA VYDSQGYATA FGPKDNPATN WANSHHRLAN LFAGIEHRFN
 201 QWKWLKAEYD Y..

15 Further work revealed the complete nucleotide sequence <SEQ ID 665>:

1 1 ATGACACGCT TCAAATATTG CCTGCTGTTT GCGCCCTGTG TGCCCCGTGA
 51 CGCGCAGCGA GATGTTCTG TTTCAGACGA CCCAAACCGG CAGGAAGCA
 101 CTGAAATTGC GACATACCC GTTACCGCGG ACCGACCCG GAGTTCCAC
 151 GACGGCTACA CTGTTTCCCGG CACCCACACC CGCGCTCGGG TGCCCATGAC
 201 CTTCGGCGA ATCCGGCGA CGCTGACCGT CAACACATCC CAACAAATGC
 251 GCGGACAAAAT CATCAAAACG CTTCGACCGGT CCTCTGGTCA GGCACGGCG
 301 ACCAGCGGCC AGATTTACGG CTTCGACCGC GCGGCTACA ACTACCTGTT
 351 CGCGCCGGG AGCCCGATCG CCAACTACCA AATCAGACCGG ATCCCCGGT
 401 CGCGGCGGT GCGGATGATC GCGAACATCCA ACACCGCGGC CTATGAGCG
 451 GTGAGAGCTG TGCGGGGGGT GCGGGGGCGT CTGGACCGCA CGGGCAGAGC
 501 TTCCGGCACCG GTCAATCTGG TCCCCAACG CCTGACCCGC AAGCCATTGT
 551 TTGAGTCTGG CGCGGAGCGG GGAACCCGA AACATTTCGG GCTGGACGG
 601 GACCTATCGG CGACCTGTAA CGCGAACGG ACCCTGGCGG GCGCCCTGGT
 651 TTCCACCTTC CGCGGCGGG ACTCTGTCGG GCGGCGGGG CGACSGCGG
 701 ATGCGAAGCT ATCAGCGACTT TTGAAATGG ACATPGCACCA CGAACACCG
 751 GTCCACCGCA CGATCGACCA AAAGAACCG CGACGCGGCC
 801 GCTACGCTG GCGCTGTACG ACAGCGAACG TTATGCCACCU GCTTGGCC
 851 CGAAAGACAA CCCCGCACCA AATTGGCGCA ACAGCGGCC CGTGGCGTC
 901 AACCTGTTG CGCGGATCGA CGACCGCTTC AACCAAAGCT GAAACACTCAA
 951 AGCGGAATGC GATACACCC CGACCGCTT CGCCAGGCC TACCGCTAG
 1001 CAGCGCTTG TTCACTGAC CACACACCG CGCGCACCGA CCTGATTCCC
 1051 GGTATTGTCG ACSCCGCACCG CGCGCACCC AGCGCCAGCC TGTCAATTGAT
 1101 CGGAAATACG CGCTGTACAC CGCGGCAACCA CGATTTAACG CGGGGTATCA
 1151 AGCGTACAA ATACCGCAGC AACAAATACG CGAACCGAG CATCATCCCC
 1201 AACCGCATTC CCAACGCTCA CGAACCTTCC CGCACGGGTG CCTACCGCGA
 1251 GCGCTCATGG TTGCGCCCAA CCATCGCGCA ATACCGCACC AGCGGCCAAA
 1301 TOGCGCGCTA TTCTGCCAAC CGTTTCCGGG CGCCCGACAA CCTTTCCCTG
 1351 ATTTCGGGC GCGATACAC CGCTTACCGC ACCGGCAGCT ACAGCACCG
 1401 CACCAACGG ATGACCTATG TTGTCGGCCTA CGCTTTACCG CCFACACAG
 1451 GCATGTTG TGCGCTGACC CGCAACCTGT CTCTTACGG CTGCTACAGC
 1501 AGCTTGTG TCCCGAACATC CGAAAAAGAC GAACACGGCA GCTACCTGAA
 1551 ACCGTTAACG GCGAACATC TGGAACCGG CGTCAAAGCC GATGCGTTG
 1601 ANGCUGCTT GAACGCTTCC CGCGCGTGT ACCGGCGCCG TAAACACAC
 1651 CGACCGACCG CAGCGAGCG ACCGGCGAGC GGCACACCTT ACTACCGCGC
 1701 CGCGAACCAA CGCAAAACCG CGCGCTGGGA ATACGAAGTC CGCGGCCGCA
 1751 CGCGCCGGA ATGCGCGATC CGCGCAGCTT ACAGCCAAAG CAAACACCGC
 1801 GACCAAGACG CGACGGCGCT GAACCCGAG AGCGTACCCG ACCGAGCTT
 1851 CAAACTTCG ACTGCTTACG ATTTGGCCC CGAACCCCCC AGCGGCTGGA
 1901 CCATGGCGC AGGGTGGGC TGCGAGCGG AAACCCACAC CGACCGTCC
 1951 AGCGTCCGGA TCCCCAACCC CGCGCCAAA GCGCCGGCGG CGACACAG
 2001 CGCGAAACCA CGCTACGGCC TGCGCACAT CTGGCGGT TACCGCTCA
 2051 ATCCGGCGGC CGAACCTGG CGAACAGCTT ACATCTGTTT CAAACACAC
 2101 TACCGCACCC AGCGCGACCC CGACAGCTAC GGCGCACITGC GGACAGTGA
 2151 CGCGCGTTT ACCTATCGGT TAAATAAA

60 This corresponds to the amino acid sequence <SEQ ID 666; ORF23-1>:

1 MTRFKYSLF RALLPVYQA DVSVSDDPKP QESTELPTIT VTADORTASSN
 51 DGTGVSTGHT PLGLMFTLRE IFOVSVISI QQMQRQNIKTT LORALLQATG
 101 TSRQIYGSDR AGYNLYFLFARG SRIANYQING IPVALADALDT GNANTAAAYER
 151 EVVRGVAGL LDGTGEPSAT VLNLVRKLRLTR KLFEVRAEAG GNRKHGFLDA
 201 DVSGSLNTEG TLRGRLVSTF GRGDSWRRRE RSRDAELYGI LEYDIAQTR

251 VHAGMDYQQA KETADAPLSY AVYDSGQYAT AFGPKDNPAT NWANSRHRAL
 301 NLFGAIEHFRF NQDWKLKAEY DTYRSRFRQPG YGVAGVLISID HNTAAFLIP
 351 GYWHDPRTH SASVSLIGKY RLFGRHEHDLI AGINGYKYAS NKYGERSIIIP
 401 NAIFNAYEFS RTGYPFPAS FAQTIPQYGT RRQIGGYLAT RFRRAODNLSL
 451 ILGGRYTRYR TGSYDSRTQC MTIVSAAETPT PTYGTIVFDLT GNLSLYGSYS
 501 SLFVPGQSQRD EHGSYSLKPVT GNNLEAGIRG EWLEGRNLNAS AAVIRAKNN
 551 LATAGRDFD GNTYYRAAQO AKTHGWEIEV GGRITPEWQI QAGYSQSXTR
 601 DQDGSRLLNPD SVVERSFSKLP TYAHFPAEAP SGWTIGACVR QOSETHHTDPA
 651 TLRIFINPAAK ARAAONSRQK AYAVADIMAR YRNPNRAELS LNVNDNLFNKH
 10 701 YRTQPDHRSY GALRTVNAAP TYREK*

Computer analysis of this amino acid sequence gave the following results:

Homology with the ferric-pseudobactin receptor PupB of *Pseudomonas putida* (accession number P38047)

ORF23 and PupB protein show 32% aa identity in 205aa overlap:

15	Orf23	6	FARGSRRIANYQINGIPVADALADTG	NANTAA	YEVVEVVRGVAGL	LDGTGEPSATVN	VLVRK	65	
	PupB	215	++GIP + I NY++GIP + L D + A +RVE+VRC	G	G	PSAT+NL+RK			
			WSRGFAIQNEYDVGVPTSTRL	DNYSQMMP	MFRVE	IVRGATGLI	SGMGNPSATIN	LIRK	
20	Orf23	66	RTRKELFEVRAAGNRKHGFLDAVGSNL	NTEXXL	RGLRVSTF	XXXXXXXX	XXXXXX	XAE	125
	PupB	274	R T + + EAGN	I G	DVG	L +	+RGR	V + +	
			RTGAEQASITGEAGNWDRYGTGFDVSGPLTETGN	NGR	FRVADYKT	EWAIDRYN	NQSQL	333	
	Orf23	126	LYGILETIAAPQTRVHAXMDYQQAKETAD	PLS	YAVD--	SQYATA	FGPKDN	PTAWHAN	183
	PupB	334	+YGI E+D++	T +	Y +	DPL	S G	T + N A +*	
25	Orf23	184	SHIRNLNFAGIEHFRNQDWKLKAEY	206					
	PupB	392	+ H + F IE +	W	K E				

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF23 shows 95.7% identity over a 211aa overlap with an ORF (ORF23a) from strain A of *N. meningitidis*:

35	orf23.pep				10	20	30	
	orf23a	QMRDQNQKIALDALLQATGTSRQIYGS	YNSYL	FARGSRRIANYQINGIPVADALADTG				
		90	100	110	120	130	140	
40	orf23.pep	NANTAA	YEVVEVVRGVAGL	LDGTGEPSATVN	VLVRK	LRKLP	FEVRAEAGNRKHGFLDAD	
	orf23a	A	LDGTGEPSATVN	VLVRK	LRKLP	FEVRAEAGNRKHGFLDAD		
		150	160	170	180	190	200	
45	orf23.pep	100	110	120	130	140	150	
	orf23a	VSGSLNTE	XLTRGRVLVSTFGRGDSWRRER	RSXAELEYCILEYDIAPQTRVHAXMDYQQAK				
		210	220	230	240	250	260	
50	orf23.pep	160	170	180	190	200	210	
	orf23a	ETADAPLSYAVYDSQYATA	FGPKDN	PTAWHAN	SHIRNL	FAGIEHFRNQDWKLKAEYD		
		270	280	290	300	310	320	
55	orf23.pep	Y						
	orf23a	Y						
		330	340	350	360	370	380	

The complete length ORF23a nucleotide sequence <SEQ ID 667> is:

1 ATGACACGGCT TCAAATATT CCGTGTGTTT GCGCCGCTGT TGCCCGTGT
 55 CGGCCAGGCC GATGTTTCG TTTCAAGCGA CCCCCAACCG CAGGGAAAGCA
 101 CTGAATTCGC GACCATCAC GTTACCGCG ACCGGACCCGC GAGTTCCAC
 151 GACGGCTACA CTGTTTCGCG CACGGCACAC CGCGCTCGGC TGCCCATGAC
 201 CCTGGCGGAA ATCCGGCAGA CGCTCAGCGT CATCACATCG CAACAAATGC
 251 GCGAACAAA CATCAAGCG CTGACCGCG CCCTTTCGA CGCGACCGGC
 301 ACCAGCGCC AGATTTACGG CTCCGACCGC GGCGCTAC AACTACCTGT
 351 CGCGCGCGGC AGCCGATCG CCAACTACCA AATCAACCGC ATCCCGCTTG
 401 CGGACCGCGT GGCGATACG GCGAATGCC ACACCGCCG CTATGAGGCC
 451 GTAGAAAGTC TTGGGGGGGT CGCCGGGGCTG CTGGACGGCA CGGGCGAGCC
 501 TTCCGCACC GTCACTTGG TGCGAACAGG CGCGACCGGC AAAGCATTGT
 551 TTGAAGTCC CGCGAACAGG GCGAACCGCA AACATTTCGG GCTGGGGCG
 601 GAGCTATCGS GCAGCTGCA TGCGGAGGC AGCGTGGGG CGCGCCCTGT
 651 TTCCACCTTG ACTCGGCGGC ACTCGGGGGT CGACGGGGGG CGCAGCGCG
 701 ATGCCGAACT CTACGGCAT TTTGAATACG ACATCGGACCC GCAAACCCCG
 751 GTCCACCGACG CTACGGACTA CGACAGGGG AAAGGAAACCC CGACGCGGCC
 801 GCTCACTACG CGCTGTGACG ACAGCGAACG TTATGCCACC GCGTGGGCC
 851 CGGAAAGACAA CCCCGACCAA AATTGGGGCA ACAGCGGCCA CGTGGGGCTC
 901 BACCTTGTG CGCGCATGC ACACCGCTTC AACCAAGACT GGAAACTCAA
 951 AGCCGGRATAC GACTACACC GAAGCGGCTG CGCGCAGGCC TACCGGGTAG
 1001 CAGCGCTCTG TTTCATCGAC CGACAAACGG CGCCCAACCG CTTGATTCCC
 1051 GGTTATTGGC AGCCGACCCG GCGCACCCAC AGCCGCAAGG TGTCATTAAT
 1101 CGGAAATAAC CGCTCTTCC GCGCCGAACA CGATTAACT GCGGPTATCA
 1151 ACGGTACAA ATACCGCAGC AACAAATACG CGAACCGCAG CATCATCCCC
 1201 AACCGCATTC CAAACGGCTA CGAATTTCGG CGCACGGGGT CCTACCGCGA
 1251 GCGTGTATCG TTGCGGACCA CGATCCCGCA ATACCGCACC AGGGCGEAAA
 1301 TCGCCGGCTG CTGCGGACCC CGAGATACAG CGGTACCGC ACCGGCAGCT
 1351 ATACTGGCGC CGAGATACAG CGGTACCGC ACCGGCAGCT AGCACAGCG
 1401 CACACAAAGG ATGACATATG TGTCGCCCAA CGCTTTACCG CCCTACACAG
 1451 GCATCTGTG CGACCTGACCG GCGAACCTGT CGCTTTACGG CTGTCACAGC
 1501 AGCTCTGTTG TCCCGCAACT GCGAAAGACG GAAACCGGCC GCTACCTGGA
 1551 ACCGGTAACC GGCACACATC TGAGACCGCG CATCAAAAGG GAATGGCTTG
 1601 AAGCCGGCTG GAAACGATCC CGCCCGGTGT ACCGGGGCGG TAAAAAACAC
 1651 CTGGCCACCG CAGCAGGACG CGACCCGAGC GCGAACACCT ACTACCGCG
 1701 CGCCAAACCA CGCAAAACCC CGCGCTGGGA AATCGAAGTC GCGGGCCGCA
 1751 TCAAGCGCCG AATCGGACATA CGAGCGAGGT ACAGCGAACAG CAAAACCCG
 1801 GACCGAGAGC CGACGGCGCT GAAACCCGAC AGCGTACCCCG AACCCAGCTT
 1851 CAAACTCTTC ACTGCTTACG ACTTTGCCG CGAAGCCCCC ACCGGCTGG
 1901 CCATCGGGCG AGCGCTGGC TGCGAGAGGG AARCCACAC CGACCCCTGCC
 1951 ACGCTCGCA TCCCAACCC CGCCGCCAA GCGCCGGCG CGACACACAG
 2001 CGCCCAAAA GCTACGCCG TCAGCGACGT CATGGCGCT TACCGCTTCA
 2051 ATCCGGGGCG CGAATCTGCG CTGAGCTGG ACAAATCTGTG CAACAAACAC
 2101 TACCGGACCC AGCCGACCG CGACACCTAC CGGGACTGC GGACAGTGA
 2151 CGCGGGCTT ACCTATGGT TAAATAAA

45 This encodes a protein having amino acid sequence <SEQ ID 668>:

1 MTRFKYSLLF AALLPVYQAQ DVSVDDPKP QESTELPTIT VTADRTASSN
 55 DGYTVSGTHT PLGLPMLTIRE IPQSWSVITS QQQRDQNIKA LDRALLQATG
 101 TSRQYIGSDR AGNYNLYFARG SRIANYING IVPADALADT GNANTAAAYER
 151 VEVVRGVAGL LDCTGPESPAT VNLVRKRPTP KPLFEVRAEA GRKHFGLGA
 201 DVSSLNAEGL TLRGRLVSTF GRGDWSWRQRE RSRDAEYLIGY LEYDIAPQTR
 251 VHAGMDYQQA KETADAPLSV AVYDSQYAT AFGPKDNPAT NWANSRHRAL
 301 NLFLAGIEHFRF NQDWKLKAEY DYTTSRSRROP YGVAGVLISID HNTAAATDLIP
 351 GYWHDADPRFLH SASVSLIGKY RLFGREHDLI AGINGYKAS NYKGEERSII
 401 NAIINPAYEFS RTGATPQFAS FAQTIPQYGT RRQIIGGYLAT RFRAADNLSL
 451 ILGGRYSRYR TGSYDSRQTC MTYVSANRFT PYTGIVFDLT GNLSSLYGSYS
 501 SLEVPQSQRD EHGSYLYKPVY GNNLIEAGIK EWLEGRNLAS AAVYRARKNN
 551 LATAAGRDEPS GNITYYRAANQ ARTHGWEIEV GGRITPEWQI QAGYSQSQKTR
 601 DQDGSRINPD SVPERSFKLRF TAYHFAEAP SWGTIGAGVR QSESTHMDPA
 651 TLRTPNPAAK ARAAONSRQF AYAVADIMAR YRFNPRABLS LNVDNLFNKH
 701 YRTQDPDRHSY GALRTVNNAF TYRFK*

ORF23a and ORF23-1 show 99.2% identity in 725 aa overlap:

	10	20	30	40	50	60	
orf23a.pep	MTRFKYSLLF AALLPVYQAQ DVSVDDPKP QESTELPTIT VTADRTASSN DGYTVSGTHT						
orf23-1	MTRFKYSLLF AALLPVYQAQ DVSVDDPKP QESTELPTIT VTADRTASSN DGYTVSGTHT	10	20	30	40	50	60

		70	80	90	100	110	120
	orf23a.pep	PLGLPMTLREI PQSVSVITSQQMRDQNIKALDRALLQATGTSRQIYGSDRAGYNYLFARG					
5	orf23-1	PLGLPMTLREI PQSVSVITSQQMRDQNIKTLDRALLQATGTSRQIYGSDRAGYNYLFARG	70	80	90	100	110
							120
		130	140	150	160	170	180
10	orf23a.pep	SRIANYQINGIPVADALADTGNNANTAAAYERVEVVRGVAGLDDGTGEPSPATVNVLVKRPTR					
	orf23-1	SRIANYQINGIPVADALADTGNNANTAAAYERVEVVRGVAGLDDGTGEPSPATVNVLVKRPTR	130	140	150	160	170
							180
		190	200	210	220	230	240
15	orf23a.pep	KPLFEVRAEAGNRKHFLGLADVSGLINNEGTLRGLVSTTFRGRDSWRQRERSRDAELYG					
	orf23-1	KPLFEVRAEAGNRKHFLGLADVSGLINNEGTLRGLVSTTFRGRDSWRQRERSRDAELYG	190	200	210	220	230
							240
20	orf23a.pep	LEYDIAPOTRVHAGMDYQQAKETADALPLSYAVYDSQGYATAFGPKDNPATNWANSRHAL	250	260	270	280	290
	orf23-1	LEYDIAPOTRVHAGMDYQQAKETADALPLSYAVYDSQGYATAFGPKDNPATNWANSRHAL	250	260	270	280	290
							300
25	orf23a.pep	NLFAGIEHRFNQDWKLKAEFYDYTRSRFRQPYGVAGVLSIDHNTAATDLIPGTWHADPRTH	310	320	330	340	350
	orf23-1	NLFAGIEHRFNQDWKLKAEFYDYTRSRFRQPYGVAGVLSIDHNTAATDLIPGTWHADPRTH	310	320	330	340	350
							360
30	orf23a.pep	SASVSLIGKYRLFGRHDLIAIGINGKYASNKYGERSIIPNAIPNAYEFRTGAYPOPAS	370	380	390	400	410
	orf23-1	SASVSLIGKYRLFGRHDLIAIGINGKYASNKYGERSIIPNAIPNAYEFRTGAYPOPAS	370	380	390	400	410
							420
35	orf23a.pep	FAQTIPOQYTRQRQIGGYLATRFRAADNLSSLGGRYTRYRTGSYDSRTQGMITYVSANRFT	430	440	450	460	470
	orf23-1	FAQTIPOQYTRQRQIGGYLATRFRAADNLSSLGGRYTRYRTGSYDSRTQGMITYVSANRFT	430	440	450	460	470
							480
40	orf23a.pep	PTYGIVFDLTGNLISLYGSYSSLLVPQSGQKDEHGSYLKPVITGNLLEAGIKGEWLEGRLNAS	490	500	510	520	530
	orf23-1	PTYGIVFDLTGNLISLYGSYSSLLVPQSGQKDEHGSYLKPVITGNLLEAGIKGEWLEGRLNAS	490	500	510	520	530
							540
45	orf23a.pep	AAVYRARKNNLATAAGRDPGSNTYYRAANQAKTHGWEIEVGGRITPEWQIQAGYSQSCTR	550	560	570	580	590
	orf23-1	AAVYRARKNNLATAAGRDPGSNTYYRAANQAKTHGWEIEVGGRITPEWQIQAGYSQSCTR	550	560	570	580	590
							600
50	orf23a.pep	DQDGSRLNPDSPVERSFKLFTAYHFAPAPSGBTIGAGVRWQSETHTDPATLRLIPNPAAK	610	620	630	640	650
	orf23-1	DQDGSRLNPDSPVERSFKLFTAYHFAPAPSGBTIGAGVRWQSETHTDPATLRLIPNPAAK	610	620	630	640	650
							660
55	orf23a.pep	ARAADNSRQKAYAVADIMARYRFNPRAELSINVDNLFNKHYRTQPDFRHSGYALRTVNAAF	670	680	690	700	710
	orf23-1	ARAADNSRQKAYAVADIMARYRFNPRAELSINVDNLFNKHYRTQPDFRHSGYALRTVNAAF	670	680	690	700	710
							720
60	orf23a.pep	TYRFKK					
	orf23-1	TYRFKK					
65	orf23a.pep						
	orf23-1						
70	orf23a.pep						
	orf23-1						

Homology with a predicted ORF from *N.gonorrhoeae*

ORF23 shows 93.4% identity over a 211aa overlap with a predicted ORF (ORF23.ng) from *N.gonorrhoeae*:

5	orf23.pep	GYNYLFARGSRIANQINGIPVADALADTGNTAAAYEREVVVRGVAGLLD	51
	orf23ng	SAVDACRIPGYNYLEFARGSRIANQINGIPVADALADTGNTAAAYEREVVVRGVAGLLD	60
10	orf23.pep	GTGEPSATVNVLVRKLRKPLFVEVRAEAGNRKHFGLADAVSGSLNTEXXLRGRLVSTFGR	111
	orf23ng	GTGEPSATVNVLVRKIPTRKPLFVEVRAEAGNRKHFGLADAVSSSLNAEGTIRGRRLVSTFGR	120
15	orf23.pep	GDSWRRRERSRXRAELYGILEYDIAPQTRVHAXMDYQQAKETADAPLSYAVYDSQGYATAF	171
	orf23ng	GDSWRQLERSRDAELYGILEYDIAPQTRVHAGMDYQQAKETADAPLSYAVYDSQGYATAF	180
20	orf23.pep	GPKDNPATWNNSRNARLNLFAGIEHHRPNQDWKLKABEDYTRSRFRQPYGVGVLISIDNS	211
	orf23ng	GPKDNPATWNNSRNARLNLFAGIEHHRPNQDWKLKABEDYTRSRFRQPYGVGVLISIDNS	240

The ORF23ng nucleotide sequence <SEQ ID 669> is predicted to encode a protein comprising

20 amino acid sequence <SEQ ID 670>:

1	SAVDACRIPG YNYLFARGSR IANYQINGIP VADALADTG NANTAAAYER
5	VVVRGVAGLLD GTGEPSATVN LVKRGRLVSTFGR LFEVRAEAGNRKHFG
10	SGLSNAEQTG RGRLVSTFGR GDSWRQLERS RDAELYGILE YDIAOPTRVH
15	AGMDYQQAKTL TADAPLSYAVYDSQGYATAF GPKDNPATWN SNSRNALNL
20	FAGIEHHRFNQ DWKLKABEDYTRSRFRQPYGVGVLISIDNS
25	WHADPRTHS SMLSMLGYRL FGREHDLLAG INGYKKYASNK YGERSIIIPNA
30	IPNAYEFQRST GAYPOPPSSA QTIPQYDTRR QIGGYLATRF RAADNLSSL
35	GGRGSYRYRAG SYNRSRQGM YVSANRFTPY TGIVFDLTGN LSLYGGSYSSL
40	FVPQLQDEH GSYLKPFTGN NSLADIKQEW LSGRLNNSA VYRARKNNLA
45	TAAQRDQSZN TYIARAQNQK THGWELEVG RITPEWIOQA GYSQSXPDRQ
50	DGSRLNPDSV PERSFLKFTA YHLAPEAPSG RTIGAGVRRQ GETHTDPAA
55	RIPNPAAKAK AVANSRKQAY AVADIMARYR FNPRTELSLN VDNLFNKHRY
60	TQPDHRSYGA LRTVNAFTY RFR*

Further work revealed the complete nucleotide sequence <SEQ ID 671>:

35	1 ATGACGACGCC TCAAACTACT CCTGCTTTTC GCGCCCTCTGC TACCCGTGTA
	51 CGCGCAGGCC GRTGTTCTGC TTTCAGACGA CCCCAAACCG CAGGAAAGCA
40	101 CCGAATTGCG RCCTCACCC AGTACCGCCCG ACCGGACACCG GAGTTCCAC
	151 GACGGCTACA CGGTTCCCG CACGGCACAC CGCTTGGCGG TGCCCATGAC
45	201 CCTGCGCGAA ATCCCGCAGA GCGTCAGCGT CATCACATCG CAACAAATGC
	251 GCGCACAAAAA CARCAAAACG CTGACACCGG CCCTGTGCA GGCGACCGGC
50	301 ACCAGCGGCC AGTTTACCGG CTCGGACGCC GCGGGCFACA ACTACCTGT
	351 CGCGCGCGGC AGCGCATCTG GCGCAATGCC ACACCGCGC ATACCGGTG
55	401 CGCACGCCCT GCGGCGATCG GCGCAATGCC ACACCGCGC CTATGAGCGC
	451 GTGAGAAGTC TCGCGCGCGT GGCGGGGGCT CGGAGCGCGA CGGGCGAGCC
60	501 TTCTGCACCA GTCACTTCTG TAGCAAAACG CCCGACCGGC AAGCCATTGT
	551 TTGAGACTCG CGCCGCAAGCC GGCAACACCCA AACATTTCGG GCTGGCGCC
65	601 GACGTATCGG GCAGCGCTGAA GCGCAAGGCC AGCGTGGCG CGCGCGCTGT
	651 TTCCACCTTC GGACGCGCGGC ACTCTGGCG GCAGCTGAA CGCAGCGCG
70	701 ATGCCAACCT ATCGCGATT TTGAAATACG ACATCGGCCG CAAACCCCG
	751 GTCCACCGAG GCATGQACTA CGACGAGGCC AAAGAAACCG CAGACGCC
75	801 GCTCACTTAC CGCCGTGTAAG ACAGCCAAAG TTATGCCACC GCTTCTGGCC
	851 CAAAAGACAA CCCCGCACA AATTTGTCGA ACAGCGCCAA CGTGGCGCTC
80	901 AACCTTTCG CCGCGATAGA ACACCCCTTC AACAAAGACT GAAACTCTAA
	951 AGCCGAAATC GACTCACCC GTAGCGCTT CGCGCACGCC FACCCTGTGG
85	1001 CAGCGCTACT TTCCATGAC CACAGCAGTC CGCGCACCGA CCTGATTC
	1051 GGTATTGTC AGCGCATCC GCGCACCCAC AGCGCGACGA TGTATGAC
90	1101 CGCGCATAC CgtctCTTC CGCGCGAGCA CSATTATTAC CGGGTATCA
	1151 ACGGCTACAA ATACGCCAGC AACAAATAG GCGAACCGAG CATATTCCC
95	1201 AACGCGCTT CCAAGCGCTA CGAATTTCGG CGCAAGGGGG CCTATTCGCA
	1251 GCGCATCATCG TTGCCCCAAA CAATCGCGCA ATACGACACC AGGGCGCAA
100	1301 TCGCGCGCTA TCTGCCAAC CGTTCCGG CGCGCGACAA CCTTTGCGTG
	1351 ATACTCGGC CGACAGATACG CGCGTACCGC CGAGCGAGCT AACAAACAGCC

1401 GACACAAGGC ATGACCTATG TGTCCGCCAA CGTTCACACC
 1451 GCATCGTTG CTGATCTGAC GSCAACCTGT CGCTTTCAGGG CTGCTACAGC
 1501 AGCTCTGTCG TCCCGCAATT GCAAAAAGRC GAAACCGGCA GTACACTGGA
 1551 ACCGCTAACG GGCAACATC TGGRAGCGCA CATCAAAGSC GAATGGCTTG
 5 1601 AAGGGCTGTG GAACCCATCC GCCCCCGCTG ACCSGGCGCG TAAAAAACAC
 1651 CTCGCCACCG CAGCAAGACG CGACCAAGAC GSCAACACCT ACTATCCSC
 1701 CGCCAACCAA GCACAAACCC ACGGCTGGGA ATTCGAAGTC GCGCCGCGA
 1751 TAAGGGCCCG ATGGCAAGATA CAGCGCAGGT ACAGCCAAAG CAACCCCGC
 1801 GACCAAGAGC CGACCGGCTG AAACCCGGAC AGCGFAcCG AACCGAGCTT
 10 1851 CAAACTTTC ACCGGCTTACCG ATCTAGGGCC CGAACGCCCG AGCGCCGCGA
 1901 CCATcggtTGC GGSTGTCGGC CGGCAAGGGG AAACCCACAC CGACCCAGCC
 1951 CGCTCTGGCA GCCTCAACCC CGCCGGCAA GCGCCGGCGC TGCCCAAACAG
 2001 CGCGCAGAAA GCCTACGGCC TCSCCGACAT CATGGCGCTG TACCGCTCA
 2051 ATTCGGCCAC CGAACTGTCG CTGAGCGTGG ACACACTGTT CRACAAACAC
 15 2101 TACCGCACCC AGCCGGACCC CGAACGCTAC GGCAGACTGC GGCGACTGAA
 2151 CGCGCGCTTT ACCTATCGGT TTAAATAA

This corresponds to the amino acid sequence <SEQ ID 672; ORF23ng-1>:

1 MTRFKYSLF ALLPVYQAQ DVSVSDDPKP QESTELPTIT VTADRTTASSN
 51 DGYPSGTHI PFLPLMTLRE IPIQSVSIVITS QQMRDQNLT KDLRALLQATG
 20 101 TSQIYGSDR AGYNLYFLARG SRIANYQING IPVADALADT GNANTAAAYER
 151 VEVVRGVAGL PDGTGEFSAT VNLYRKHPTR KPLFEVRAEAA GRKHFGLGA
 201 DVSGSLNAEG TLRGRLVSTF GRGDWSRQLRE RSRDAELYGI LEYDIAPOTR
 25 251 VHAGMDYQQA KETADPLSY AVYDSQGYAT AFSPKDNPAT NWNSNRNRAL
 301 NLFGAGIEHFRH NQDWKLKAET DYTRSFRQF YVGAVGLSID HSTAATDLIP
 351 GYWHADPRH SASSMLTQKYL RLFGREHDLI AGINGYKAS NKYGERSIIP
 401 NAIPNAYEFS RTGAYPOQSF FAQIPOYDT RRQIGGYLAT RFRADNLSL
 451 ILGRGRYSR YAGSYNSRTQG MTYVSANRFT PYTGIVFDLT GNLSLYGSYS
 501 SLFPVQLQKD EHGSYLKPFV GNNLEADING EWLEGRLNAS AAVYRARKNN
 551 LATAAQRDQS GTNTYVRAANO AKTHGWEIEV GGRITPEWQI QAGYSOSKPR
 30 601 DQDGSRLNPD SVPERSFKLFV TAYHLAEAP SGRTGIGAVR RQGETHTDPA
 651 ALRPINPAAK ARAVANSRQK AYAVADIMAR YRFNIPRTELS LNVDNLFNKH
 701 YRTQPDHRHSY GALRTVNAAF TYRFK*

ORF23ng-1 and ORF23-1 show 95.9% identity in 725 aa overlap:

	10	20	30	40	50	60
35 orf23-1.pep	MTRFKYSLF ALLPVYQAQ DVSVSDDPKP QESTELPTIT VTADRTTASSN					
orf23ng-1	MTRFKYSLF ALLPVYQAQ DVSVSDDPKP QESTELPTIT VTADRTTASSN					
	10	20	30	40	50	60
40 orf23-1.pep	PLGLPMTLRE IPIQSVSIVITS QQMRDQNLT KDLRALLQATG TSQIYGSDRAGYNYFLARG					
orf23ng-1	PFGLPMTLRE IPIQSVSIVITS QQMRDQNLT KDLRALLQATG TSQIYGSDRAGYNYFLARG					
	70	80	90	100	110	120
45 orf23-1.pep	SRIANYQINGIPVADALADTGNANTAAAYERVEVVRGVAGLLDGTGEPSATVNLRKHLTR					
orf23ng-1	SRIANYQINGIPVADALADTGNANTAAAYERVEVVRGVAGLLDGTGEPSATVNLRKHLTR					
	130	140	150	160	170	180
50 orf23-1.pep	KPLFEVRAEAGNRKHFGLDADVSGSLNTEGTLRGRLVSTFGRGDSWRRERSDAELYGI					
orf23ng-1	KPLFEVRAEAGNRKHFGLDADVSGSLNTEGTLRGRLVSTFGRGDSWRRERSDAELYGI					
	190	200	210	220	230	240
55 orf23-1.pep	LEYDIAPOTRVHAGMDYQQA KETADPLSY AVYDSQGYATAFGEKDNPATNWNSRNRAL					
orf23ng-1	LEYDIAPOTRVHAGMDYQQA KETADPLSY AVYDSQGYATAFGEKDNPATNWNSRNRAL					
	250	260	270	280	290	300
60 orf23-1.pep	NLFAGIEHFRNODWKLAEYDYTRSFRQPYGVAGVLSIDHNTAATDLIPGYWHADPRH					
orf23ng-1	NLFAGIEHFRNODWKLAEYDYTRSFRQPYGVAGVLSIDHNTAATDLIPGYWHADPRH					
	310	320	330	340	350	360

		310	320	330	340	350	360	
5	orf23-1.pep	370	380	390	400	410	420	
	orf23ng-1	SASVSLIGKYRLFGRHDLIAGINGKYKAVSNKYGERSTIPNAIPNAYEFVRTGAYPOPAS SASMSLTGKYRLFGRHDLIAGINGKYKAVSNKYGERSTIPNAIPNAYEFVRTGAYPOPAS	370	380	390	400	410	420
10	orf23-1.pep	430	440	450	460	470	480	
	orf23ng-1	FAQTIPQYGTRRQIGGYLATRFFRAADNLSSLILGGGRYTRYRTGSYDSRTQQMTYVSANRFT FAQTIPQYDTRQIGGYLATRFFRAADNLSSLILGGGRYTRYRTGSYDSRTQQMTYVSANRFT	430	440	450	460	470	480
15	orf23-1.pep	490	500	510	520	530	540	
	orf23ng-1	PYTGVIFDLTGNLNSLYGGSYSSLFVPQSGKDEHGSYLSKPFVGNLEAGIKGEWLEGRLNAS PYTGVIFDLTGNLNSLYGGSYSSLFVPQLGKDEHGSYLSKPFVGNLEAGIKGEWLEGRLNAS	490	500	510	520	530	540
20	orf23-1.pep	550	560	570	580	590	600	
	orf23ng-1	AAVYRARKNNLATAAGRDPQSGNTYYRAANQAKTHGWIEVGGRTIPEWQIQAGYSQSCKTR AAVYRARKNNLATAAGRDPQSGNTYYRAANQAKTHGWIEVGGRTIPEWQIQAGYSQSCKTR	550	560	570	580	590	600
25	orf23-1.pep	610	620	630	640	650	660	
	orf23ng-1	DQDGSRSLNPDSVERSKFLTFAYHFAPAPSGWTIGAVGRWQSETHDPTALRIPNPAAK DQDGSRSLNPDSVERSKFLTFAYHFAPAPSGRTIGAGVVRQGETHTDPAALRIPNPAAK	610	620	630	640	650	660
30	orf23-1.pep	670	680	690	700	710	720	
	orf23ng-1	ARAADNSRQKAYAVADIMARYRFNPRALSLSNVNDLFNKHYRTQPDFRSYSGALRTVNAAP ARAVANSRQKAYAVADIMARYRFNPRTELSLSNVNDLFNKHYRTQPDFRSYSGALRTVNAAP	670	680	690	700	710	720
35	orf23-1.pep	770	780	790	800	810	820	
	orf23ng-1	TYRFKK TYRFKK	770	780	790	800	810	820
40	orf23-1.pep	TYRFKK TYRFKK	770	780	790	800	810	820
	orf23ng-1	TYRFKK TYRFKK	770	780	790	800	810	820

In addition, ORF23ng-1 shows significant homology with an OMP from *E.coli*:

45	sp P16869 FHU E COLI OUTER-MEMBRANE RECEPTOR FOR FE(III)-COPROGEN, FE(III)-FERRIOXAMINE B AND FE(III)-RHODOTRULIC ACID PRECURSOR	>gi 1651542 gnl PID d1015403 (D90745) Outer membrane protein FhuE precursor [Escherichia coli] >gi 1651545 gnl PID d1015405 (D90746) Outer membrane protein FhuE precursor [Escherichia coli] >gi 1787344 (AE002010) outer-membrane receptor for Fe(III)-coprogen, Fe(III)-ferrioxamine B and Fe(III)-rhodotrulic acid precursor [Escherichia coli] Length = 729 Score = 332 bits (843), Expect = 3e-90 Identities = 228/717 (31%), Positives = 350/717 (48%), Gaps = 60/717 (8%)
55	Query: 38 TITVTADRTAASSN--DGYTVSGTHTPPGLPMLTREIPOSVSVIITSQQMDQNINKTLDRAL 95 T+ V - TA + + Y+V+ T + + MT R+IPQS+ + + Q-M DQ ++ TL + Sbjct: 43 TVIVEGSATAPDGENDYSVTSAGTQMQRDIFQSVTIVSQRMEDQQLQTGEVM 102	
60	Query: 96 LQATGTTSRQIYGSDRAGYNYLFARGSRRIANYQINGIP-----VADALADTGNNATAA 147 G S+ SDRA Y + +RG + I NY + +GIP + DAL+D A Sbjct: 103 ENTLGISKSKQADSRALY--YSRGFDIINDYMVQGIPTYFESRWNLGDALSM----AL 154	
65	Query: 148 YERVEVVRGVAGLPDGTDGEFSATVNLRVKHPTKPLF-EVRAEAQRKUFGLGADVSGSL 206 +EREVVRVG GL GTG FSA -+NRVKH T + +V AE G+ Ad+ L Sbjct: 155 FERVEVVRGATLGMGTGNFSAAINMRKHATSRFKGDSVAEYGSWNKERYVADLQSLP 214	
70	Query: 207 NAEGTLLRGLRVLSTFGRGDSWRQLERSRDAELYGILEYDIAPOTRVHAGMDYQQAKETADA 266 +G +R R+V + DSW S GI++ D+ T + AG +YQ+ + Sbjct: 215 TEDGKIRARIVGYYQNNDNSWLDRYNSEKTFSGIVDADLGDTTLSAGYEYQRIDVNSPT 274	
	Query: 267 PLSYAVYDSQGYATAFGPKDNPATNWNSNSRNRA1NLFAGIEHRFNQDWKLKAEGDYTRS 326	

+++ G + ++ + A +W+ + +F + + F W+ ++

Sbjct: 275 WGGGLPRWNTDGSSNSYDRARSTAPDWAYNDEINKVFMILKQQFADTWQATLNATHSE 334

Query: 327 F -RQPQGVAGVLSDIDHSTA--TDLIPGY-----WHADPRTHSA-SMSLTGKYRLFG 374
F + Y A V D + + PG+ W++ R A + G Y LGF

Sbjct: 335 FDISKMMVVDAVYNKADGMLVLPGVPSNYGPFGFDIVGGTQWNSGRKRVDALDLFADGSYELFG 394

Query: 375 REHDILAGINGYKYASNKYGER--SIIPNPAIYNAYEFSRTGAYPQPSSFAQTIPQYDTRR 432
R++L+ G Y +NAY +I P+ I + Y F+ G +PC Q++ Q DT

Sbjct: 395 RQRNLMFG-GSYSKQNRRYFSSWANIPFDEIGSFYNNF--GNFCTDWSFPQSLSAQDITTHE 451

Query: 433 QIGGGYLATRRAAIDLSSLILGGRSYRAGGSNSRTQGMTY-VSANRFTPYTGYIVFDXXX 491
Y ATR AD L LILG RY+ +R + +TY + N TPF G+VF

Sbjct: 452 MKSLYVAATRVLADPLHLILGARYTNWRVDT-----LTYSMEKNHTTPYAGLVFDIND 504

Query: 492 XXXXXXXXXXXXXXXXXPCOLKDEHGSQLLPKVNNEADATKGEWLEGRINASAAVYRARKNNL 551
F PQ + D + YL P+TGN E +K +W+ RL + A++R + +N+

Sbjct: 505 NWSTASYASITSIFQPQNDRDSSGKYLAPITGNYYELGLKSDWWNNSRLLTTLAIFRIEQDNV 564

Query: 552 ATAAAGR---DQSGNTYRAANGQAKTHOWELEVGGRITPEWQIQAGYSQSKPDRDGDGSRLN 608
A + G +C T X+A + + G E E + G IT WG+ G ++ D +G + H +N

Sbjct: 565 AQSTGTPIPGSNGSETAYKAVDGTVSKGVFELNAGITDNWQLTGATRYIAEDNEGRANV 624

Query: 609 PDSPVERPSFKLFTYHLAEPAPSGRTGIVAGVRRQGETHDPAAIRIPNPAAKARAVANSR 668
P + +P + K+FT+Y L P P +T+G V Q +TD P RA

Sbjct: 625 P-NELRTTVRMFTTSYRL-FVMPE-LTVGGGVNWQNRVYDTV-----TPYGTFR----E 672

Query: 669 QKAYAVADIMARYRFNPNRTIELSLANVDNLNFNKHYRTQPDFRH-SYGALRTVNAAFTYRF 724
Q +Y+A+ D+ RY+ L NV+NLF+K Y T + YG R + TY+F

Sbjct: 673 QGSYALVLDLTRYQVTNFSLQGNVNNLFDKTYD TNVEGSIVYGTPRNFSITGTYQF 729

Based on this analysis, it was predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF23-1 (77.5kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 15A shows the results of affinity purification of the His-fusion protein, and Figure 15B shows the results of expression of the GST-fusion in *E.coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for Western blot (Figure 15C) and for ELISA (positive result). These experiments confirm that ORF23-1 is a surface-exposed protein, and that it is a useful immunogen.

40 Example 80

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 673>:

```

1   ATGCGCACGG CAGTGGTTT GCTGTGATGC ATGCCGATGG CGGCTTCGTC
51   GCGCATGATG CCGGAATGG TGTGCGCGG CGTGTGCGCG CGAACCGGCAA
101  TCATATCCAA GCGCACCAA AAAACGCGG TCATGCGCTC GAGTTTGTCC
151  AGCGTCAGcA CGCGCTCTTC GGCGgcGgca ATCATACCTT CGCTCTCGGA
201  AACGGGGATA AACG:GCCAC TCAACACCC GACCGCGCTG GAAGCCATCA
251  TGCGGCCTT TTTCAGCGCA TCGTGTAGCA ATGCCAAAGG TGCTGTGTTG
301  CGTGGCTAC CGCAGACGCT CAAGGCCCCATT ThITCAAGAG TGCGTGCAC
351  ThAGTGGCG ACAGGG..
```

50 This corresponds to the amino acid sequence <SEQ ID 674; ORF24>:

```

1   MRTAVVLLLI MPMAAASSAMM PEMVCAGVSP GTAIISKPKTE QTAVMASSLS
51   SVSTPASAAA IIIPSSSETGTI NAPLKPPTAL EAIMPFFTA SFNSNAKAAVV
101  PCVPQTLKPI XSRMRATXSP TG..
```

Further work revealed the complete nucleotide sequence <SEQ ID 675>:

```

      1 ATGGCGCACGG CAGTGGTTT GCTGTTGATC ATGCCGATGG CGGCTTCGTC
      5 GGCAATGTG CCGGAAATGG TGTCCGGGG CGTGTGCGG GAAACGGCAA
     10 TCATATCCA GCGGACCSAA CAAACCGGGC TCATGCGCTC GAGTTTGCC
     15 AGCTCTAGGC CGCGCTGCTC GCGCGGGCGA ATCATACCTT CTGCTTCGGA
    20 AACGGGATAAACGGGAC TCAAAACCCC GACCGCGCTC GAAGGCATCA
    25 TGGCGGCTT TTTCACCGGA CGTGTAGCGA ATGCCAAAGC TGCTGTGTTG
    30 CGTGCCTA CGCACAGCCT CAAGCCCAT TCTTCAGAA CGCGTGCAC
    35 TGAGTCCGGC ACGGCGGGG CGGGGGCGAC CGACAACTCG AGAATACCAA
   10 ACGGGATATT CAGCATTTT GAGGCTTCG GCGCGATGAG TTGGCCCAAG
   15 CGGGTAATTG TGAAGACGAT TTTCCTCACT ACTTCGGCAA CTTCGGTCAA
   20 TGTCGTTGCA TTGATGATTT CAAACSCGCG TTTTACGACA CTGGGCGCG
   25 ATAGCGGAC ATTGATACG GACATCGCGT CGCCCGAACCG ATGAAACGGG
   30 CCCGCATATAACCGGGTGTG TTCCACCGGG TTGCGAACAA CGAACATTIT
   35 AGCCGCGGAAACCCCTCGG GCGGTGATTG CGCGCGGGT TTGACGCGTT
   40 CGCCGCCAACTTGTG CTCATATTGA TACCGGGACCG CGTAGCTGCG
   45 ATATGATGAG AGCTGACAC AATATCGTGA GTCTTCATCG CTTCGGAAAT
   50 GGAGCGGATT AACACCTCAT CGGAGGGCGA CATCCCCCTT TGACCAAACG
   55 CGGGATAAAAAG GACACCGCA TGCGTTTGCG AGCTTATATCC
   60 901 AAAGTTGGG CCACGCTGAC GTAA

```

This corresponds to the amino acid sequence <SEQ ID 676; ORF24-1>:

```

      1 MRTAVVLLI MPMAASSANM PEMUCAGVSP GТАIISKPKTE QTAVMASSLS
      5 SVSTPASAAA IIPSSSETGI NAPLKPTL EAIMPPFFTA SFNSAKAAVV
     10 PCVEOTLKP I SSMRMATESP TAGVGASDKS RIENGFISIF EASRPMSSPT
     15 RVIKRAVFTT TSATSVNVA SEFSNAFFT PGPDPTPLIT ASASPEF*NA
     20 PAINGLSTA LQNTTILAQP KPSGVISAVR LTVPASLTA SILIPARVLP
     25 ILMELHTISV VFIASGMERI NTSSSEGDIPE CTNAEKPPIK DIPMALAALS
     30 KVCAITL*

```

Computer analysis of this amino acid sequence gave the following results:

30 **Homology with a predicted ORF from *N.meningitidis* (strain A)**

ORF24 shows 96.4% identity over a 307 aa overlap with an ORF (ORF24a) from strain A of *N.meningitidis*:

	10	20	30	40	50	60	
35 orf24a.pep	MRTAVVLLIIMPMAASSAMMP EMVCAVGSPGTAIISXPTSQTAVIASSLNVSTPASAAA						
orf24	MRTAVVLLIIMPMAASSAMMP EMVCAVGSPGTAIISXPTSQTAVMASSLSVVSTPASAAA	10	20	30	40	50	60
40 orf24a.pep	IIPSSSXGTGINAPLKPTALEAIMPFFTASFSNAKAAVVPCVPQTILKPISSRMATESP	70	80	90	100	110	120
orf24	IIPSSSETGGINAPLKPTALEAIMPFFTASFSNAKAAVVPCVPQTILKPISSRMATESP	70	80	90	100	110	120
45 orf24a.pep	TAGVGASDKSRIPNGIFSI FEASRPMSSPTRVILKAVFETTSATSVNVVASEFSNAFFT	130	140	150	160	170	180
orf24	TAGVGASDKSRIPNGIFSI FEASRPMSSPTRVILKAVFETTSATSVNVVASEFSNAFFT	130	140	150	160	170	180
50 orf24a.pep	PGDPTPTLITASASPEPXNAFAIXGLSSXALQNTTILAQPKPSSVISXVRMLVSPASLTA	190	200	210	220	230	240
orf24	PGDPTPTLITASASPEPXNAFAINGLSSXALQNTTILAQPKPSSGISVAVRLTVSPASLTA	190	200	210	220	230	240
55 orf24a.pep	SILIPARVLPILMEHTISVVFIAASGMERINTSSEGDIPECTSAEKPPIKDTPMALAALS	250	260	270	280	290	300
orf24	SILIPARVLPILMEHTISVVFIAASGMERINTSSEGDIPECTNAEKPPIKDTPMALAALS	250	260	270	280	290	300

orf24a.pep	KVCATLTX
orf24	KVCATLTX

- 5 The complete length ORF24a nucleotide sequence <SEQ ID 677> is:

	1	ATGGCCGACGG	CAGTGGTTT	GCTGTGTTGATC	ATGGCCGATTCG	CGGCCGTGCG
	51	GGCAATGATG	CGGAAATTGG	TGCTGGCGGC	TGTCGCGG	GGGAGGCAGGAA
10	101	TCTATATCCA	NCCGACCGAA	CAACCGGGC	TCTATCCGCTTC	GAGTTTATCCG
	151	AACGTGACGA	GGCTGCTTC	GGCGCCGCGA	ATCATACCTT	CTCTTCGGG
	201	NAACGGGATA	AACGGGCGAC	TCAACAGCCC	AAACGGGCTC	GAAGGCTCATCG
	251	TGGCCGGCTT	TTCATCAGGA	TGCTTGACGA	ATGCGAAAGC	TGCTGTGTTG
	301	CTGGCGGTAC	CGCGACGCT	CAANACCCAT	TCTTCAAGAA	TGCGGCGGCA
	351	CGAGGTGGCC	ACGGCGGAC	CGACGGGCG	CGACGGATCG	AGAACATCCAA
15	401	ACGGGATATG	CAGCATTTTG	GAGGCTCTGC	GGCGGATAG	TTCTGGCACAG
	451	CGGGTTAATT	TGAAAGGGG	TTCTTCACAA	ATCTTCGGG	CTTCGGCTGCG
	501	TGTCGTGTC	TCCGAATT	CCAAAGGGCG	TTTACGACA	CCGGGGCGCG
	551	ATATAGCCGAC	ATTAATACCA	GCATCTGGCTT	CGCTCTGGCC	TAAGGACGAGC
	601	CCGGCCATAT	ACGGGTTTC	TCCTCCGGCC	TCTCAGAACAA	CGACGGATTTC
	651	GGGGCGACGG	AAAACCTTCA	TGTGATTTC	ANCGCTGCGT	TATGTTGTTT
20	701	CGCCCCCGAG	TCTGACGGCC	TCTTATTTGA	TAATGGCGGC	CGTACTGGCC
	751	ATATGGATGG	AGCTGGCAC	GATATCTAGTA	CTCTTCATCG	CTTCGGGAAAT
	801	GGAAACGGAT	AAACACTCTG	CAGAAAGGCA	TAACATTTCT	TGCAACAGGCG
	851	CGGGAAATGG	GGCAATTAATA	GACACCCCGA	TGGCTTTGGC	AGCTCTTATCC
	901	AAATGTTGCG	CCACGGCTGA	GTAA		

- 25 This encodes a protein having amino acid sequence <SEQ ID 678>

```

1 MRTAVVLLLI MPMMAASSMM PEMVCAGVSP GTAIISXPTE QTAVIASSLS
51 NVSTPESAAA IPISSXXTQI NAPLKPPTAL BAIMPPFFTA SFSNAKAAVW
101 PCFPVQLKEI SSRMRATESP TAGVGASAKS RIFNGFISF EASRMRMSSPT
151 RYVILKAVFFF TSATSVNYY SEFSNAAFTP GPGDFTPLIT ASASPE*P*
201 FAIXGLLSXA LQNTTILAQP KPPVSIVSRX LMWSRSPASITA SILIPARVLE
251 ILMEHLTIVS FVIASGMERX NTSSEGDIPE CTSAEKKPIK DTPMALAALS
301 KVCATIT*

```

It should be noted that this protein includes a stop codon at position 198.

ORF24a and ORF24-1 show 96.4% identity in 307 aa overlap:

		10	20	30	40	50	60
35	orf24a.pep	MRTAVLVLLIMPEMASSAMMPEMVCAGVSPGTAIISXPTEQTAIVIASSLSNVSTPASAAA					
	orf24-1	MRTAVLVLLIMPEMASSAMMPEMVCAGVSPGTAIISXPTEQTAIVIASSLSNVSTPASAAA					
40	orf24a.pep		10	20	30	40	60
	orf24-1		70	80	90	100	120
45	orf24a.pep	IIPSSSXTCINAPLKPTTALEAIMPPFFTASFSNAAKAVVPCVPQLKPISSRMRATESP					
	orf24-1	IIPSSSETCINAPLKPTTALEAIMPPFFTASFSNAAKAVVPCVPQLKPISSRMRATESP	70	80	90	100	120
50	orf24a.pep		130	140	150	160	180
	orf24-1	TAGVGASDKSRIPNGIFSIFEASRPMSPTTRVILKAVFETTSATSVNVVASEFSNAFTT					
55	orf24a.pep		130	140	150	160	180
	orf24-1	PGPDPTPTLITASASPEPXNAPAIKGSLXALQNTTILAQPKPSSVISXVRLMVSPLSTA	190	200	210	220	240
60	orf24a.pep	PGPDPTPTLITASASPEPXNAPAIKGSLXALQNTTILAQPKPSSVISXVRLMVSPLSTA	190	200	210	220	240
	orf24-1	SILIPARVLPILMELHTISVFIASGMRKNTSSEGDIIFCTSAAEKPPIKDTPMALAALS	250	260	270	280	300

	orf24a.pep	KVCATLTX
5	orf24-1	KVCATLTX

Homology with a predicted ORF from *N.gonorrhoeae*

ORF24 shows 96.7% identity over a 121 aa overlap with a predicted ORF (ORF24ng) from *N.gonorrhoeae*:

10	orf24.pep	MRTAVVLLILIMPMAASSAMMPPEMVCAGVSPGTAIISKPTEQTAVMASSLSSVSTPASAAA	60
	orf24ng	MRTAVVLLILIMPMAASSAMMPPEMVCAGVSPGTAIMSKPTEQTAVMASSLSSVNTPASAAA	60
15	orf24.pep	IIPSSSETGINAPLKPTTALEAIMEPPFTASFSNAKAAVVPCCVPQTLKPISSRMRATESP	120
	orf24ng	IIPSSSETGINAPLKPTTALEAIMEPPFTASFSNAKAAVVPCCVPQTLKPISSRMRATESP	120
	orf24.pep	TG	122
20	orf24ng	!: TAGVGASDKSRMFNGIFSIFEASRPMSSTPDRVILKAVFFTSATSVRLTASEFSSAALT	180

The complete length ORF24ng nucleotide sequence <SEQ ID 679> is:

```

1 ATGCCCACGG CGGGTTTTT GCTGTTGATC ATGCCGATGG CGGCTTCGTC
51 GGGCATGATG CGGGAAATGG TTGCGGCAGG CTGTCGGCCG GGAAACGGCAA
25 101 TCATGTCACCA ACCAACCGAG GAGACCGGGG CTATGGCTTC GAGTTGGTCC
151 AGCGGCAAC CGCGCTTCGTC GCGGGGGCGA ATCATACATT CTGCTTCGGGA
201 AACGGGGATA AACGGCCCGC TCAACACCCC GACCGGGCTG GAAGGCATCA
251 TGCGGCCCTT TTTCACCGGA TCGCTCAGCA ATGGCCAAAGC TGTTGTTG
301 CGCTGGTAC CGCGAGAGG TCAACGGGTT TCTTCAGAAA TGCGGGCCAC
351 CGAGTGGCG ACAGGGGGGGG TCGCTGGCAG CGACAAATCC AGATAGTCCGA
30 401 ACAGGGTATT CRGCATTTTG GAGGCTTCGC GRCCGATGG TTGGCCACG
451 CGGGGTATTG TGAAGGGGGT TTTCCTTACG ACTTCGGCGA CTCGGTCA
501 GCTGACCCGG CGCGGATTTT CCAGCGCGC TTTCGACAGC CTCGACCCGG
551 ATACCGCGAC ATAATACACA ATTCATCGCTT CGCCCGAGCC GTGGAACGCA
601 CCCGGCATAC AGCGGATTCG TTCCACCGGG TTGCGAGACAA CGACGATTTT
651 GGGCGAGCG AACACTTCGG GTGTGATTTG AGCCGTCGCT TTGATGTTT
701 CGCGCTGCCG CTTGACCGCA TCCATATTGC TACCCGGAGG CGTGTGCGC
751 ATATTCATGG AGCTTCACAC GATATCGGTA GTTTTCATCG TTTCGGGAC
801 GGACCGGATC AACACTCTT CGCGAGGGCA CATACTTTT TGACCAAGCG
851 CGGAAAGCC CGCCGATAAAG GACACCCGA TGGCTTGGC TGCCCTTGCC
40 901 AAAGCTCGG CAACGCTGAC ATA

```

This encodes a protein having amino acid sequence <SEQ ID 680>:

```

1 MRTAVVLLILIMPMAASSAMMPPEMVCAGVSPGTAIISKPTEQTAVMASSLSSVSTPASAAA
51 SVNTPSAAA IIPSSSETGAI NAPLKPTTALEAIMEPPFTASFSNAKAAVVPCCVPQTLKPISSRMRATESP
45 101 FCVPQTLKPI SSRMRATESP TAGVGASDKS RMPNGIPSIF EASRPMSST
151 RVILKAVFTT SEFSAAALTT PGPDTPTLIT ASASFPWPNA
201 PAINGLSSTA LQNNTTIAQPK KPGVVISAVR LMVSPASLTA SILIPARVLP
251 ILMELHTIVF VFIASGTERI NTSSEGDIPF CTSAEKPKI DTPMALAALS
301 KVCATLT*

```

ORF24ng and ORF24-1 show 96.1% identity in 307 aa overlap:

50	orf24-1.pep	10 20 30 40 50 60
	orf24ng	MRTAVVLLILIMPMAASSAMMPPEMVCAGVSPGTAIISKPTEQTAVMASSLSSVNTPASAAA
55		10 20 30 40 50 60
	orf24-1.pep	IIPSSSETGINAPLKPTTALEAIMEPPFTASFSNAKAAVVPCCVPQTLKPISSRMRATESP
60	orf24ng	IIPSSSETGINAPLKPTTALEAIMEPPFTASFSNAKAAVVPCCVPQTLKPISSRMRATESP

-390-

		130	140	150	160	170	180
	orf24-1.pep	TAGVGASDKSRMNGIFSIPEASRPMSSPTRVILKAVFETTSATSVNVVAEFSNAAFT					
	orf24ng	TAGVGASDKSRMNGIFSIPEASRPMSSPTRVILKAVFETTSATSVRLTASEPSSAALT					
5		130	140	150	160	170	180
	orf24-1.pep	190	200	210	220	230	240
10	orf24ng	PGPDPTTLITASASPEPXNAPAINGLSSTALQNTTILAQPKPSPGVISAVRLTVSPASLTA					
	orf24-1.pep	PGPDPTTLITASASPEPXNAPAINGLSSTALQNTTILAQPKPSPGVISAVRLTVSPASLTA					
15	orf24ng	190	200	210	220	230	240
	orf24-1.pep	250	260	270	280	290	300
	orf24ng	SILIPARVLPIMLELHTISVVFIAASMERINTSSEGDIPECTNAEKPPIKDTPMALAALS					
		250	260	270	280	290	300
20	orf24-1.pep	KVCATLTX					
	orf24ng	KVCATLTX					

Based on this analysis, including the presence of a putative leader sequence (first 18 aa – double-underlined) and putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 81

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 681>:

30	1	..ACCGACGTGC	AAAAAGAGTT	GGTCGGCGRA	CAACGCAAGT	GGGCGCAGGA	
	51	AAAATCATGC	RACTGCCGAC	AAAGCCGCCG	GCAGGCGACG	CGGCAGGAAT	
	101	ACGCCGATAA	CCTCAAGCTG	CAATGCGACA	CGCGGATGAC	GGCGGAAACGG	
	151	ATACAGTATC	TTGGCGCTA	TTCCATCGAT	TAG		

This corresponds to the amino acid sequence <SEQ ID 682; ORF25>:

35	1	..TDVQKELVGE	QRKWAQEKKIS	NCRQAAAQAD	RQEYREYLKL	QCDTRMTRER	
	51	IQYLRGYSID	*				

Further work revealed the complete nucleotide sequence <SEQ ID 683>:

40	1	ATGTATCGGA	AACCTATTCG	GCTGCCGTTT	GCCCTGCTGC	TTGCCCGCTTG	
	51	CGCGCAGGGAA	GAACGCCCA	AGGCATTGGA	ATGCCCAAC	CCGGCCGTGT	
	101	TGCAAGCAT	ACCGCGCAAT	ATTTCAGAAA	CGCTCACGCA	GGAAAGCGCT	
	151	TCTTGTGGC	CGCGAACGCG	CAACCGGTTT	GTCGATGCCG	ACAAANTTAT	
	201	CGCGCCGCC	TACGGTTGG	GGTTTCTTT	GGAACACGCT	TGGAAACACG	
	251	AGGAAGGGCG	GGCGACCTTC	TGTATGCCG	ATTGAAACAT	TACCGTGGCG	
	301	TCTGAAACGG	TTGGCGATGC	CAAGGAAAC	AGCCCCCTGT	TGTACGGGGA	
	351	AACTGTTTC	TGGATATTCG	TGGCGCAGAA	GACGGGGGGC	ATATGCGAGT	
	401	TTAAAGACGG	CGTATTAGCG	GCAGCGCTG	GCTTCCTGCC	CGTCAAAGAC	
	451	GGTCAGACGG	CATTGTCGA	CAACACGGT	GTTATGCCG	GGCAAACACGT	
	501	GTCCTGGCCG	CTGCTGCCCT	ACGGCGTGA	GAGCATCGT	ATGATAGACG	
	551	GCAAGGGCGT	GAAGAAAGAA	GAAGCGCTCA	GGATTTTGAG	GGAAAAGACG	
	601	CTGTAAGGAG	AACCTCGCAA	ACCCACGCC	GAAGACATT	TGGAACACAA	
	651	TGCCGGCCCG	GGCGATCCG	GGCTACCCCA	AGCCGCAGAA	GGCCGCCCG	
	701	AACCGGAAAT	CCTGCTACGT	GAAGCGGGG	AGCGTGGCGA	TACCGTTACC	
	751	GTATCACGGG	GGGAAGATGGA	AGAGGGGGC	GTACAAAACCC	AGCGTGGCGA	
	801	ATCCGAATT	ACCAACATT	GGGGAGACT	CGTACCGAC	GTGCAAAAG	
	851	AGTTGTCGG	CGAACAAACG	AAAGGGCGC	AGGAAAATAAT	CAGCAACTGC	
	901	CGACAAAGCCG	CGGGCGAGG	AGACCGGGAG	GAATACGCCG	AATACCTCAA	
	951	GCTGCAATGC	GACACGGCA	TGACGCGCGA	ACGGATACAG	TATCTTCGCG	
	1001	GCTATTCAT	CGATTAG				

This corresponds to the amino acid sequence <SEQ ID 684; ORF25-1>:

```

 1 MYRKLIALLPF ALLLLAACGRE EPPKALECAN PAVLQGIRGN IQETLTQEARN
 51 SFAREDGRQF VDADKIIAAR XXXXXSLEHRA SETQEGGRFT CIADLNITVTP
 101 SETLADAKAN SPLLYGETAL SDIVRQQTGG NVEFKDGVLT AAVERFLPVKD
 151 GQTAFVINTV GMAAQTLSAAR LLPPVGKSVI MIDGKAVKRE DAVRILSGKA
 201 REEEPSKPTP EDILEHNNAAG GDAGVPOAAE GAPEPEILHP DDGERADTVT
 251 VSRGVEEAR VQNQRAESEI TKLWGLLDTD VQKELVGEQR KWAQEKSNC
 301 RQAAAQADRQ EYAEYLKLQC DTRMTRERIQ YLRGYSID*

```

Computer analysis of this amino acid sequence gave the following results:

10 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF25 shows 98.3% identity over a 60aa overlap with an ORF (ORF25a) from strain A of *N. meningitidis*:

		10	20	30
15	orf25-pep	TDVKQELVGEORKWAQEKSNCRQAAQAD		
	orf25a	VTVSRRGEVEEARVQNQRAESEITKLWGLLDTDVQKELVGEEXRKWAQEKSNCRQAAQAD	250 260 270 280 290 300	
20	orf25-pep	40 50 60		
	orf25a	RQEYAEYLKLQCDTRMTRERIQYLRGYSIDX RQEYAYEYLKLQCDTRMTRERIQYLRGYSIDX	310 320 330	

The complete length ORF25a nucleotide sequence <SEQ ID 685> is:

```

25      1 ATGATATCGGA AACTCAATTGC GTGTCGTTT GCCCCGTGTC TTGCGCTTG
 51 CGGAGGGA GAAAGCCCCA AGGCATTGGA ATGCCAACAC CCCCGCTGT
101 TGCAGAATG AGCCGCAAT ATTCAAGGAA CGCTCAAGGCA GGAAAGCGCT
151 TCTTCGCGC GCGAAAGCNG CANCAGCTT GTGCGATGCC ACNAAATTAT
201 CGCCGGCGCC TANGNNNN NGNTNTCTTT GGAAACAGCTG TCGGAAACCG
251 AGGAAGGCCG GCGCACGTC TGTNTCGCG ATTGAACAT TACCGTGCCTG
301 TCTGAAACGC TGCGCAATGC AAAGGCCAAC AGCCCCCTGC TGACCGGGGA
351 AACCGCTTGC TGGATGATATG TCGCGCAAGGA GACGGCCGCG AATGTCGAGT
401 TTAAAGACGG CGTATGAGC GCAGCCGCTC GTCTCTTACG CCTCAAAGAC
451 GGTCAAGANG CTTTGTGGA CAACACGGTC GGTATGGCG CGCAACAGCT
501 GTCTGGCGG TTGCTGCTT ACAGCGTGRA GAGCATCGTG ATGATAGACG
551 GCAAGGGGT AAAAAGAAGA GACCGCGGCA GGATNTGAG CNGANAAGCC
601 CGTGAANAAAG AACGGCTTCA ANCCNGCCC GAAAGACATT TTGGAACATTA
651 TGCGCCGGG GGGGATGAGC AGCTAACCCA AGCCGGAGAA GACGCCGGCG
701 AACCGAAAT CCTGCATCT GACGACGGCG AGCGTGCCTG TACCGTACCC
751 GTATGCTGGG GCGAAATGGA AGAGGGCCCN GTACAAAACAG AGCGTGGGGA
801 ATCCGAATT CCTAACATTG GGGGGAGACT CGATACCGAC GTGCAAAAG
851 AGTGGTGC GGAAANACGG AGATGGGCC AGAAAAAAAT CAGCAACTTG
901 CGAACAGCGG CGGCCGAGGG AGACCGGCCG GAATACGCC AATACCTCAA
951 GCTGCAATGC GACAGCGCGA TGACGGCGA ACGGATACAG TATCTCGCG
45 1001 GCTATTCATG CGATTAG

```

This encodes a protein having amino acid sequence <SEQ ID 686>:

```

 1 MYRKLIALLPF ALLLLAACGRE EPPKALECAN PAVLQXIRXN IQETLTQEARN
 51 SFAREDGXQF VDADKIIAAR XXXXXSLEHRA SETQEGGRFT CIADLNITVTP
 101 SETLADAKAN SPLLYGETAL SDIVRQQTGG NVEFKDGVLT AAVERFLPVKD
 151 GQXAFAVINTV GMAAQTLSAAR LLPPVGKSVI MIDGKAVKRE DAVRILSGKA
 201 REEEPSKXXP EDILEHNNAAG GDAGVPOAAE GAPEPEILHP DDGERADTVT
 251 VSRGVEEAR VQNQRAESEI TKLWGLLDTD VQKELVGEQR KWAQEKSNC
 301 RQAAAQADRQ EYAEYLKLQC DTRMTRERIQ YLRGYSID*

```

ORF25a and ORF25-1 show 93.5% identity in 338 aa overlap:

55	orf25a.pep	10 20 30 40 50 60
	orf25-1	MYRKLIALLPFALLLAACGREGPPKALECANPAVLQGIRGNIQETLTQEARSFAREDGXQF

		10	20	30	40	50	60
5	orf25a.pep	70	80	90	100	110	120
	orf25-1	VDADXIIIAAXXXXSLEHASETQEGGRTFCXADLNITVFSSETLADAKANSPLLYGETAL 70	80	90	100	110	120
10	orf25a.pep	130	140	150	160	170	180
	orf25-1	SDIVRQK7GGNVEFKDGVLTAARFLPVKDGQXA FVNDTVGMAAQTL SALL PYGVKSIV SDIVRQK7GGNVEFKDGVLTAARFLPVKDGQXA FVNDTVGMAAQTL SALL PYGVKSIV 130	140	150	160	170	180
15	orf25a.pep	190	200	210	220	230	240
	orf25-1	MIDGKAVKKEDAVRIXSXKXAREXP SKXXFEDILEHNAAAGGDADVPQAGDAPEPEIHLH MIDGKAVKKEDAVRILSGKAREEEPSKPTFEDILEHNAAAGGDAGV PQAEGAPEPEIHLH 190	200	210	220	230	240
20	orf25a.pep	250	260	270	280	290	300
	orf25-1	DDGERADTVTSRGEVEEARVNQNRASEEITKLWGGLDT DVQKELVGEXRKWAQE KISNC DDGERADTVTSRGEVEEARVNQNRASEEITKLWGGLDT DVQKELVGEXRKWAQE KISNC 250	260	270	280	290	300
25	orf25a.pep	310	320	330	339		
	orf25-1	RQAAAQADRQEYAEYLKLQCDTRMTTRERIQYLRGYSIDX RQAAAQADRQEYAEYLKLQCDTRMTTRERIQYLRGYSIDX 310	320	330			
30	orf25-1						

Homology with a predicted ORF from *N.gonorrhoeae*

ORF25 shows 100% identity over a 60aa overlap with a predicted ORF (ORF25ng) from

35 *N.gonorrhoeae*:

	orf25.pep	TDVQKELVGEQRKWAQE KISNC RQAAAQAD	30
	orf25ng	VTVSRRGEVEEARVNQNRASEEITKLWGGLDT DVQKELVGEXRKWAQE KISNC RQAAAQAD	308
40	orf25.pep	RQEYAEYLKLQCDTRMTTRERIQYLRGYSID	60
	orf25ng	RQEYAEYLKLQCDTRMTTRERIQYLRGYSID	338

The complete length ORF25ng nucleotide sequence <SEQ ID 687> is:

45	1	ATGTACATCGGA AACTCATTTG GCTGCCGTTT GCCCCCTGTC TTGCGACGCTG	
	51	CGGAGGAGGA GAA CGCCCAA AGGGT GTGA ATGCCAAC CCCCGCTGT	
	101	TGCAGGACAT ACCCGGCA GT ATT CAGGAAA CGCTCACCGCA GGAAAGCCGCT	
	151	TCTTTCGGCC GCGAAGACGG CAGCGCTT GTCGATGCGC ACAAAATTAT	
	201	COCGCCGCC TACGGTTTG CTGTTCTTGT GGAAACAGCT TC GAAAAGCC	
	251	AGGAAGGGCG GCAGCACGTC TTATGCCG ATT TGACAT TACCGTGC	
	301	TCTGAAACGG TTGCCATGTC CGAGGCAAC AGCCCCCTGC TTATGGGG	
	351	AACCTTTTG CGACAGATCG TGCGCAGAGA GACGGGGCGC AATGTCGAGT	
	401	TTAAAGACGG CGTATTGACG GCAGCGCTC GCTTCCTGCC CGCCAAAAGAC	
	451	GCTGGACGG CTTTATCGA CAACACGGTC GGTATGGCGA CGCAAACGCT	
	501	GTCTGCCGC TTGCTGCCCTT AC GCGCTGAA GCGATCCTG ATGATAGACG	
	551	CGAAGGGCGT GACAAGAAA GACCGCGTCA GGGTTTTGAG CGCCAAAAGCC	
	601	CGTGAAGAAG AACCCCTCAA ACCCCCCC GAAAGCATT TTGAAACACA	
	651	TGCCGCCGCC GCGGATGCCGCG GCGTACCCCA AGCCCGCAGAA GCGCCACCCG	
	701	AACCCGAAT CCTGCATCC GACAGCTGC AGCGTGC CG TA CGCTTAC	
	751	GTATCACGGC CGCAATGGA AGAAGGGCGC GTACAAAAC ACATGTCGGGA	
60	801	ATCCGAATT ACCAACATTT GGGGGAGCT CGATACCGAC GTGCAAAAG	
	851	AGTTGGTCCG CGAAGACGGC AA GTGGGGCC AGGAAAAAAAT CAGcaactgc	
	901	cGCAAGACGG CGCGCGAGGC AGACCGGCAG GAATACGCCG AATACCTCAA	
	951	GCTCCATGCG ACACCGCGGA TGACCGCGA ACggatTACAG TATCTTCGCG	
	1001	GCTATTCAT CGATTTAG	

65 This encodes a protein having amino acid sequence <SEQ ID 688>:

1 MYRKLIALLP ELLLAACGREG EPPKALECAN PAVLQDIRGS IQETLTQEARM
 51 SFAREDGRQF VDADKIIIAAA YGLAFLSLEHA SETQEGGRRTF CIADINITVTP
 101 SETLDAEAN SPILLYGETSL ADIVQQKTTG NVEFKDGQVLT AAVRFLPLPAKD
 151 ARTAFIDNTV GMATOTLSAA LLPYGVKSIV MIDGKAVITKE DAVRVLSGKA
 201 RBEEPSPKPTP EDILEHNNAAG GDAGVEQRAE GAPERELILHP DOVERADTVT
 251 VSRCGEVEBAR VQNQRAESEI TKLWGGLDTD VQRELVGEQR KWAQEKTISNC
 301 RQAAAQADROQ EYAEYIQLQCDTRMTERIQYLRGYSID*

ORF25ng and ORF25-1 show 95.9% identity in 338 aa overlap:

		10	20	30	40	50	60
10	orf25-1.pep	MYRKLIALLP FALLLAACGREG EPPKALECAN PAVLQDIRGS IQETLTQEARM					
	orf25ng	SFAREDGRQF VDADKIIIAAA YGLAFLSLEHA SETQEGGRRTF CIADINITVTP					
		10	20	30	40	50	60
15	orf25-1.pep	70	80	90	100	110	120
	orf25ng	VDAOKIIIAAYGLAFLSLEHASETQEGGRRTFCIADLNITVPESETLADAKANSPLLGYGETAL					
		70	80	90	100	110	120
20	orf25-1.pep	130	140	150	160	170	180
	orf25ng	SDIVRQKTTGGNVEFKDGVLTAAVRFLPVKDGQTAFVNDTGVMAAQTLSAALLPYGVKSIV					
		130	140	150	160	170	180
25	orf25-1.pep	190	200	210	220	230	240
	orf25ng	MIDGKAVTKEDAVRVLSGKAREEEPSKPTPEDILEHNAAAGGDAGVPQAEGAPEPEILHP					
		190	200	210	220	230	240
30	orf25-1.pep	250	260	270	280	290	300
	orf25ng	DDGERADTVTVSRGEVEEARVNQRAESEITKLWGGLDTDVQRELVGEQRKWAQEKTISNC					
		250	260	270	280	290	300
35	orf25-1.pep	310	320	330	339		
	orf25ng	RQAAAQADROQ EYAEYIQLQCDTRMTERIQYLRGYSIDX					
		310	320	330			
40	orf25-1.pep						
	orf25ng	RQAAAQADROQ EYAEYIQLQCDTRMTERIQYLRGYSIDX					
		310	320	330			

45 Based on this analysis, including the presence of a predicted prokaryotic membrane lipoprotein lipid attachment site (underlined) in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF25-1 (37kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described 50 above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 16A shows the results of affinity purification of the GST-fusion protein, and Figure 16B shows the results of expression of the His-fusion in *E.coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for Western blot (Figure 16C), ELISA (positive result), and FACS analysis (Figure 16D). These experiments confirm that ORF25-1 is a surface-exposed protein, and 55 that it is a useful immunogen.

Figure 16E shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF25-1.

Example 82

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 689>

```

5   1 ATGCAGCTGA TCGACTATT ACATTCAATT TTCTCGGTG TGCCACCCCT
    51 TTGGCAGCTG GCACCTTCGG TCATTAACCGG CGCGCTACIG CTGTCCTTAA
1001 101 GCATCGGTAT TCTGWySGC GTTGGCTTTTG TTGGTGGGGG CAACCCCGTC
151 151 GACGGCTGCA CACACTGAA AGACATGGTC TGCGCTTGG CTGGTGCAGA
201 201 CGSYGATTGG TGCGTGGGCA AACAAAATACT TTGGTTTTTC CKGATACCTT
251 251 TGGGTTTTT TACTTCCTG CTGACCTTAC CGGCAGCAA T.....
```

//

```

10 851 ..... AC TTTCGCTGGTA
901 901 TTTCGGGGCA CTGGGGCGGT CTTCGGCGTC GTTCTCTGCA CGTCGGCAC
951 951 GATTAACACC CGCCGACTATC CAAAGACCGT TTGGCAGGGT CGAAATCTA
1001 1001 TGTTGGGGCA AATTCGCAATT TGATTCCTCG CTGGCTCAT CAGTACGGT
1501 1501 GTCGGGGAAAC TGCAACCGG CGATTACACTT CCACACCTGG TTGGGGCAA
1101 1101 CATCACCTCC GCTTCCTTCG CGCTCATCCCTT CTTCCTGCTC GGCAAGGGTA
1151 1151 TGGCGTTTGC CACAGGGACA AGCTGGGGAA CGTTGGCGAT TATGCTGGCG
1201 1201 ATTGGCGGCC CCAATGGGGT AGACATGGCA CCCGGCTGTGA TTATCCCGTG
1251 1251 TATGTCGGCA ATAATGGCGG GGGCGGTATG CGGGCACCC TGTCGGCCA
20 1301 1301 TTTCGGCACAA GACCATTCCTG TGCTTACCCG GGCGGCCCTG CAACACACATC
1351 1351 GACCAAGTTA CCTTCGCAACT GCTTCAACCG TTAAACCGTT CGCGCCCGC
1401 1401 CGCATGGGGT TACCTCCGAT TGGGTCGAC AAAATCCGGG CTGTTGGGCT
1451 1451 TTGGCAGGAC AGGGATTGTA TTGGCGGTGC TGATTTTTCT GTTGAAGAT
1501 1501 AAAAAA..
```

25 This corresponds to the amino acid sequence <SEQ ID 690; ORF26>:

```

1 MQLIDYSHF FSVVPPFLAL ALAVITRRVL LSLIGIILXX VAFLVGGNPV
51 DGLTHLKDMMV VGLAWSKDXW SLGKPKILVF XILLGIFTSL LTYSGSN...  

      //
```

```

30 251 ..... TSLW
301 301 FGGTCGVFAV VLCTLGFIKT ADYPKAVWQG AKSMFGAIIA LILAWLISLV
351 351 VGMEMHTGDLR STLVAGNIIHP GFLPVILFL AVSMAMAFATGT SWGTFGIMLP
401 401 IAAAMAVKVE PALIIPCMSA VMAGAVGCDH CSPISDTIL SSTGARCNHI
451 451 DHVTSQLPYA LTVAAAAASG YLALGLTKSA LLGFGTGIV LAVLIFLLKD
501 501 KK..
```

35 Further work revealed the complete nucleotide sequence <SEQ ID 691>:

```

1 ATGCAGCTGA TCGACTATT ACATTCAATT TTCTCGGTG TGCCACCCCT
51 TTGGCAGCTG GCACCTTCGG TCATTAACCGG CGCGCTACIG CTGTCCTTAA
1001 101 GCATCGGTAT TCTGWySGC GTTGGCTTTTG TTGGTGGGGG CAACCCCGTC
151 151 GACGGCTGCA CACACTGAA AGACATGGTC TGCGCTTGG CTGGTGCAGA
201 201 CGSYGATTGG TGCGTGGGCA AACAAAATACT TTGGTTTTTC CTGATACCTT
251 251 TGGGTTTTT TACTTCCTG CTGACCTTAC CGGCAGCAA TCGGGGTTT
301 301 GCGCAACCGCA CATTAAACCGA TAACTCTTAT CGACGACTAT TCCCCACAGTC
351 351 GACCGCTGCA CTGGCTTGGC TAACCTTAT CGACGACTAT TCCCCACAGTC
401 401 TCGGGCTGGC TGCGATGGCC CGGGCGGTGA CGGACAAAGT TAAAGTTTCC
451 451 CGCACCAAAAC TGCGCTTACG CTCGACTCTC ACTGCGCTC CTATGCGGT
501 501 GCTGAGTGGC TGCGTCAAGT CGGGCGGTGA GATTATCGCC ACGGTGGCG
551 551 GACTGCTGCT TACCTCAAAA ATCACCGAT CACACCGGAT GGGGAGGTTT
601 601 GTGCGCCATGAA GCGCTGATGAA CTACATGGCA CTGTTTGCCC TGATTATGGT
651 651 GTTGGCCTGC CGATGGTTTG CTCTGGCATC CGCTCGCATG CGACGTTTGG
701 701 AACAGCGGC GTTGAACGAA GCCACAGATG AAACCTGGCT TTCAAGACGGT
751 751 ACCAACAGTC GTGTTTACCG ACTGATTAT CGCGTTTTTG CCTTAATCGC
801 801 CTCACAGGTT TCGCGCATGA TCTACACCGG CGCGCAGGCA AGCGAAACCT
851 851 TCAGCAATTG GGGGGCTATT GAAACACGG AGCTAAACAC TTGGCTGGTA
901 901 TTTCGGGGCA CTGGGGCGT CTGGCGCTG TTCTCTGCA CGCTCGGAC
951 951 GATTAACACC CGGCAGCTGGT CAAAGACCGT TTGGCAGGGT CGAAATCTA
1001 1001 TGTTCGGCGC AATTCGCAATT TAAATCTCGT CTGGCTCAT CAGTACGGT
1051 1051 GTCGGGGAAAC TGCAACACGG CGATTAACCTC TCCACACTGG TTGGGGGCAA
1101 1101 CATCACCTCC CGCTTCCTGC CGCTCATCTT CTTCCTGCTC CGCGAGGTGA
1151 1151 TGGCGTTTG CGACAGCTGGC AGCTGGGGGA CGTGGCGCAT TATGCTGGCG
1201 1201 ATTGGCGGCC CCAATGGCGT CAAAGCGAA CGCGCGCTGA TTATCCCGTG
1251 1251 TATGTCGGCA GTAAATGGCGG GGGCGGTATG CGGCAGGACAC TGCTCGGCA
1301 1301 TTTCGGACAC GACCATTCCTG TGTCACCGG CGCGCGCGT CGACACATC
```

1351 GACACAGGTTA CCTCGCAACT GCCTTACGGC TTAACCGTTG CGCGCGCCGC
 1401 CGCATCGGGC TACCTCGCAT TGGGTCTGAC AAAATCGGG CTGTTGGGCT
 1451 TTGGCACGAC AGGCATGTGA TTGGCGGTGC TGATTTTCT GTTGAAGAT
 1501 AAAAAACCGG CCAACGCCCTG A

5 This corresponds to the amino acid sequence <SEQ ID 692; ORF26-1>:

1 MOLIDYSHSF FSVPFPFLAL ALAVITRRVL LSLGIGILVG VAFLVGGNPV
 5 DGLTHLKDMDV VGLAWSDDGR SLGPKLILV LLLLGIFTSL LYSGSNQAF
 10 101 ADWAKRHINK RRGAKMLTAC LVEVTFFIDY FHSLAVGAI A RPTDKEFKVS
 151 RTKYLAILDS TAAPMCVLM P VSSWGSASIA TLAGLLVTVK IITEYPTPMGF
 201 VAMELMNYYA LFALIMVFV AFWSEDIISM ARFEQAAARNE AHDETAVASDA
 251 TKGRVYALI FVLAIIASV SAMIYTGAOA SETFSILGAF ENTDWNTSLV
 301 FGGTCGVYLAV VLCTLTGTTK ADYTPKAWOG AKSMFGAAIAI LILAWLISTV
 351 VGEMLITGDSL STLVAGNCIHP GFPVPLFLL ASVMAFATGT SWTFGIMLP
 401 IAAAMAVKVE PALLIPCMSA VMAGAVCGDH CSPISDTTIL SSTGARCNHI
 451 DIVTSOLPYA LTVAAAAASG YIALGLTKSA LLGFGTTGIV LAVLIFLLKD
 501 KGRKA*

Computer analysis of this amino acid sequence gave the following results:

Homology with the hypothetical transmembrane protein HI1586 of *H.influenzae* (accession number P44263)

ORF26 and HI1586 show 53% and 49% amino acid identity in 97 and 221 aa overlap at the

20 N-terminus and C-terminus, respectively:

Orf26	1	MOLIDYSHSFVSVPPFLALALAVITRRVLSSLGIGILVG VAFLVGGNPV	DGLTHLKDMDV	60
	5	M+D+S S +S+VP LA+ LA+ TRRV	L +L V	
HI1586	14	MELIDFSSVWSIVPALLAIILAIATRVRVLSLSSAGIISLMLSDWQIGSAFNLYLVKNV	73	
Orf26	61	VGLAWSDXDWSLGLPKHFLVXILLGIFTSLTYSGSN	97	
	5	V L ++D + + I+F +LLG+ T+LLT SGSN		
HI1586	74	VSLVYADGEIN-SNNMIVLFLLLGLVLTALLTVSGSN	109	
Orf26	86	IFTSLLTYSGS--NTSLVFGGTGVFAVVLCTL--GTTIKTADYPKAVWQGAKSMFGXXXX	141	
	5	+F+ L T + TS LVG GG C + L + + +Y ++ G KSM G		
HI1586	299	VFSVLGTFENNTVVGTSVLVGGFCSSIISTILLDRQSVPEVRSWIVGIKSMGAAIAI	358	
Orf26	142	XXXXXXXXSTVVGEMHTGYDLSTLVAGNIHPPGFLPVILFLLASVMAFATGTSWTFGIMLP	201	
	5	+ +VGM+ TG YLS+LV+GNI FLPVIL+L + MAFT+GTTSWTFGIMLP		
HI1586	359	LFFPAWTINVKIVGDMQTMGTGKYLLSVSNIPOMFLPVILFVGLAANAAFSTGTTSWTFGIMLP	418	
Orf26	202	IAAAAMAVKVEPALLIPCMSAVMAGAVCGDHCSPISDTTLISSTGARNHIDHVTSQXXXX	261	
	5	IAAAAMA P L+ +PC+S+MAVGAVCGDHCS+SDTTLISSTGARNHIDHVTSQXXXX		
HI1586	419	IAAAAMAAAANAAPELLLPCLSAVMAGAVCGDHCSPISDTTLISSTGAKCNHIDHVTTQLPYA	478	
Orf26	262	XXXXXXXXXXXXXXKSALLGGFGTGTIVLVLAVLIFLLKD	302	
	5	S L GF T + L V+IF +K +		
HI1586	479	ATVATATSIGYIVVGFYISLAGFAATAVSLIVIIFAVKKR	519	

Homology with a predicted ORF from *N.meningitidis* (strain A)

Orf26 shows 58.2% identity over a 502aa overlap with an ORF (ORF26a) from strain A of *N.meningitidis*:

50	orf26.pep	10 20 30 40 50 60	MOLIDYSHSFVSVPPFLALALAVITRRVLSSLGIGILVG VAFLVGGNPV
	orf26a	10 20 30 40 50 60	DGLTHLKDMDV
55	orf26.pep	70 80 90 99	MOLIDYSHSFVSVPPFLALALAVITRRVLSSLGIGILVG VAFLVGGNPV
	orf26a	70 80 90 100 110 120	DGLTHLKDMDV
60	orf26.pep	70 80 90 99	VGLAWSDXDWSLGPKHLVFXILLGIFTSLTYSGSNXX-----
	orf26a	70 80 90 100 110 120	VGLAWSDXDWSLGPKHLVFXILLGIFTSLTYSGSNQAFADWAKRHINKRNRRGAKMLTAC

	orf26.pep	-----
5	orf26a	<u>LVFVTFIDDYFHS</u> LAVGAXARPVTDFKFVSRKLAYILDSTAAPMCVLMPVSSWGASIIA 130 140 150 160 170 180
10	orf26.pep	-----
	orf26a	<u>TLAGLLVITYKITEYTPMGT</u> FVAMSMLNNYYALFALIMVFFVVAWFSEPIGSMARFEQAALE 190 200 210 220 230 240
15	orf26.pep	----- 100 110 TSLV
	orf26a	AHDETAVS <u>DGSWGRVY</u> ALIIPVLALIASTVSAMIYTGQAQESTFSILGAFENTDVNTSLV 250 260 270 280 290 300
20	orf26.pep	----- 120 130 140 150 160 170 EGGTCGVFAVVILCTLGTIADYPKAWQGAKSMFGATAIILILAWLISTVUGEMHTGDYL : : orf26a <u>EGGTCGV</u> IAVVILCTLGTIADYPKAWQGAKSMFGATAIILILAWLISTVUGEMHTGDYL 310 320 330 340 350 360
25	orf26.pep	----- 180 190 200 210 220 230 STLVAGN <u>IHPGFLPV</u> ILFLLASVMFA <u>TGTSWGT</u> FGIMPLPIAAAMAVKVEPA <u>LIIPCMSA</u> : : orf26a <u>STLVAGN</u> IHPGFLXVILE <u>LLASVMFA</u> TGTSWGT <u>FGIMLPIAAAMAVKVDPSL</u> LIIPCMSA 370 380 390 400 410 420
30	orf26.pep	----- 240 250 260 270 280 290 VMAGAV <u>CGDHCSPI</u> DTTILS <u>STSGARCNH</u> IDHVTSQL <u>PYALT</u> VAAAASGYIALGLTKSA : : orf26a <u>VMAGAV</u> CGDHCSPI <u>DTTILS</u> STSGARCNH <u>IDHVTSQL</u> PYALT <u>VAAAASGYIALGLTKSA</u> 430 440 450 460 470 480
35	orf26.pep	----- 300 310 LLGFRTGIVLAVL <u>FLLKDKKK</u> : : orf26a <u>LLGFXTGIVLAVL</u> FLLKDKKRANAX 490 500
40	orf26.pep	----- 300 310 LLGFRTGIVLAVL <u>FLLKDKKK</u> : : orf26a <u>LLGFXTGIVLAVL</u> FLLKDKKRANAX 490 500

The complete length ORF26a nucleotide sequence <SEQ ID 693> is:

45	1	ATGCAGCTGA TCGACTATT CACATTCAATT TTCTCGGTTG TGCCACCCCTT
	51	TTTGCACTGA CGCACTGCCG TCATTACCGG CGCGCTACTG CTGCTTTAG
	101	GCACTCGTAT TCTGCTGCCG TTGCTCCCTA TTGGTGGCCGG CAACCCCGTC
	151	GACGCTCTGA CACACCTGGA GTTACATGGTC GTCGGCTTGG CTGGCTCAGA
	201	CGGCGATTG TGCGCTGGCA ACCCCAAANT CTGGTTTTTC CTGATACATT
	251	TGGGATTTTT TACTCCCTG CTGACCTRACT CGCGCAGCAA TCAAGGGTTT
	301	GCGCTCTGGT CAAACCGGCA CATTAAAGGAA CGCCGCGCCGG CGAAAATGCT
	351	GACGCCCTGA CTGGCTGTTGC TAACCTTAT CGACGACTAT TTCCACAGTC
	401	TGCGCTGGG TGCGCTGCG CGGCCCGTTA CGGACAAGTT TAAAGTTTCC
	451	CGGCCCTAACAC TCGCTCATAT CTCGCACTCC ACTGGCGCGC CTATGTGCGT
	501	GCTGATGGCC GTTCAAGCTT GGGGGCGCTG GATTATCGCC ACGCTTGGCG
	551	GACTGCTGTC TACCTCACAA ATCACCGGAT ACACGCCGAT GGGGAGCTTT
	601	GTCGGCCATGA GGCTGATGAA CTATACGGA CTTGGTGGCC TGATATTGTT
	651	GTTGCGTGTG GCATGGTTCCT CTTGGACGAT CGCGCTGATG GCACGTTTCG
	701	AAACAGCCG TGTGACGCAA GCCCGAGAT AAACCTGGCTT TTACAGACGG
	751	AGCTGGGGCA GGTTTAACGC ATTGATTATAT CCCGTTTTGG CCTTAATCGC
	801	CTCACCGGT TCGCGCATGA TCTACACGG TGACAGGGCA AGCGAAAACCT
	851	TCAGCACTTATT GGCTGCTCATTT GAAATACGG ACGCTGAACAC TTGGCTGGTA
	901	TTCGGGGGCA CTGGCGCTG GCTTGGCGTC GTCTCTGCA CGCTCGGCAC
	951	GATTAAATAC CGCGATATAC CAAAGGCGT TTGGCAGGGT CGGAATATCCA
	1001	TGTTCCGGCA AATCCCGCA TTAATCCTTG CCTGGCTCAT CGATCACGGTT
	1051	GTCGGCGAAA TGCAACACAGG CGACATACCTC TCCAGCTGTC TTGGGGCCAA
	1101	CATCCATCCC GGCTCTCTGN CGCTCATCCT TTCTCTGCTC CGCACGGCTGA
	1151	TGGGCGTTTG CGACAGGCAAC AGCTGGGGGA CGTTOGGCGAT CATGCTGCG
	1201	ATTGCGCGGC CCATGGCGGT CAAAGGCGAT CCTCTACTGA TTATCCCGTG
	1251	TATGTCGCCG TGATGGCGG GGGCGGTATG CGGCACAC TGCCTCGCCA
70	1301	TTCCGACAC GACCACCTTG TCGTCCACCG GGGCGCGCTG CAACACACATC

```

1351 GACCAAGCTTA CNTCGCAACT GCCTTACGCC TTAACCGTTG CGCCCGCCGC
1401 CGCATGGGN TACCTCCCAT TGGGTCTGAC AAAATCGGGG CTGTTGGGT
1451 TTGGCANGAC AGGCATTGTA TTGGCGGTGC TGATTTCT GTTGAAAGAT
1501 AAAAAACGCG CCAACGCGT A

```

5 This encodes a protein having amino acid sequence <SEQ ID 694>:

```

1 MQLIDYSHSF FSVVPPFLAL ALAVITRRVL LSLGIGILVG VAFLVGGNPV
5 DGLTHLKDMD VGLAWSGDW SLGKPKXLVLF LLLGIFTSL LTYSGSNQAF
10 ADWAKRHKIK RRGAKMLTAC LVFVTFFIDDY FISLAVGAXA RPVTDKFKVS
151 RAKLAYILDS TAAPCMVLMP VSWSGASIIA TIAGLILTVK IITEYTPMGTF
201 VAMSILNYYA LFALIMVFVV AWFSFDIGSA ARFEQAALINE AHDETAVSDG
251 SWGRVYALII PVLAJIASTV SAMIYHGQA SETFSILGF ENTDVNTSLV
301 FGGTTCGVILAV VLCLTGITKI ADYPKAWQG AKSMEGAIAAI LILAWLISIV
351 VGEMLHTGDYL STLVAGNIPF GFLXVILFLM ASWMAFATGT SWGTFGIMLP
401 TAAAMAVKVD PSLLIIPCMSA VMAGAVCGDH CSPFISDTIL SSTGARNCHHI
451 DHVTSQLPYA LTVAAAAASG YLALGLTKSA LLGFGXTGIV LAVLIFLKD
501 KKRNAA*

```

ORF26a and ORF26-1 show 97.8% identity in 506 aa overlap:

		10	20	30	40	50	60
20	orf26a.pep	MQLIDYSHSFSSVVPFLALALAVITRRVLSSLGIGILVGVAFLVGGNPV	DGLTHLKDMD				
	orf26-1	MQLIDYSHSFSSVVPFLALALAVITRRVLSSLGIGILVGVAFLVGGNPV	DGLTHLKDMD				
25	orf26a.pep	VGLAWSGDWSLGKPKXLVFLILLGIFTSLLTYSGSNQAFADWAKRH	IKNNRGAKMLTAC				
	orf26-1	VGLAWSGDWSLGKPKILVFLILLGIFTSLLTYSGSNQAFADWAKRH	IKNNRGAKMLTAC				
30	orf26a.pep	LVEFTFIDDYFHSLAVGAXARPVTDKFVSRALKAYILDSTAAPMCV	IMLPVSSWGASIIA				
	orf26-1	LVEFTFIDDYFHSLAVGATARPVTDKFVSRALKAYILDSTAAPMCV	IMLPVSSWGASIIA				
35	orf26a.pep	130 140 150 160 170 180					
	orf26-1	130 140 150 160 170 180					
40	orf26a.pep	TLAGLILTVKYKITEYTPMGT	FVAMSILNYYALFALIMV	VVAVWFSDIGSMARFEQAALINE			
	orf26-1	TLAGLILTVKYKITEYTPMGT	FVAMSILNYYALFALIMV	VVAVWFSDIGSMARFEQAALINE			
45	orf26a.pep	190 200 210 220 230 240					
	orf26-1	190 200 210 220 230 240					
50	orf26a.pep	AHDETAVSDGSGWGRVYALII	PFLALIASTVSAMIYTGAQ	QASETSFSI	LGFENTDVNTSLV		
	orf26-1	AHDETAVSDGSGWGRVYALII	PFLALIASTVSAMIYTGAQ	QASETSFSI	LGFENTDVNTSLV		
55	orf26a.pep	250 260 270 280 290 300					
	orf26-1	250 260 270 280 290 300					
60	orf26a.pep	310 320 330 340 350 360					
	orf26-1	310 320 330 340 350 360					
65	orf26a.pep	370 380 390 400 410 420					
	orf26-1	370 380 390 400 410 420					
	orf26a.pep	STLVAGNIPHGFVLVFLILASVM	FAFATGTSWGTGIMPLIAAA	AMAVKVVDPSLI	IPCMSC		
	orf26-1	STLVAGNIPHGFVLVFLILASVM	FAFATGTSWGTGIMPLIAAA	AMAVKVVEPALI	IPCMSC		
	orf26a.pep	370 380 390 400 410 420					
	orf26-1	370 380 390 400 410 420					
	orf26a.pep	430 440 450 460 470 480					
	orf26-1	430 440 450 460 470 480					
	orf26a.pep	VMAGAVCGDHCSPISD	TTLSTS	GRCNHIDHVTSQLPYALT	VAAAASGYLALGLTKSA		
	orf26-1	VMAGAVCGDHCSPISD	TTLSTS	GRCNHIDHVTSQLPYALT	VAAAASGYLALGLTKSA		
	orf26a.pep	430 440 450 460 470 480					
	orf26a.pep	490 500					
	orf26a.pep	LLGFGXTGIVLAVLIFL	LLKDKKKRANAX				

orf26-1	: : : : LLGFGTTGIVLAVLIFPLLKDKRKRANAX 490	500
---------	---	-----

5 Homology with a predicted ORF from *N. gonorrhoeae*

ORF26 shows 94.8% and 99% identity in 97 and 206 aa overlap at the N-terminus and C-terminus, respectively, with a predicted ORF (ORF26ng) from *N. gonorrhoeae*:

10	orf26.pep	MQLIDYSHSFFSVVPPFLALALAVITRRVLLSLGIGILXXVAFLVGGNPVDGLTHLKDVM	60
	orf26ng	MQLIDYSHSFFSVVPPFLALALAVITRRVLLSLGIGILVGVAFLVGGNPVDGLTHLKDVM	60
	orf26.pep	VGLAWSDXDWSLGKPKILVFXILLGIFTSLTYSGSN	97
15	orf26ng	VGLAWADGDWSLGKPKILVFLILLGIFTSLTYSGSNQAFADWAKRHIKNRCGAKMLTAC	120
		/	
20	orf26.pep	TSLVFGGTGCGVFAVVLCTLGTTIKTADYPKA	326
	orf26ng	ASTVSAMITYGAQASETSILGAFENTDVNTSLVFGGTGCGVLAVALCPTIKTADYPKA	326
	orf26.pep	VWQGAKSMFGAIAILLILAWLISTVVGEMHTGDLVTLVAGNTHPGFLPVILPFLLASVMAF	386
25	orf26ng	VWQGAKSMFGAIAILLILAWLISTVVGEMHTGDLVTLVAGNTHPGFLPVILPFLLASVMAF	386
	orf26.pep	ATGTSWGTGFGIMPLPIAAAMAVKVEPALIIPCSMASVAGAVCGDHCSPISDTTLSTGAR	446
	orf26ng	ATGTSWGTGFGIMLPPIAAAMAVKVEPALIIPCSMASVAGAVCGDHCSPISDTTLSTGAR	446
30	orf26.pep	CNHIHDVTSQLPYALTVAAAAAGYLAGLTKSALLFGFTGIVLAVLIFLKKDKKK	502
	orf26ng	CNHIHDVTSQLPYALTVAAAAAGYLAGLTKSALLFGFTGIVLAVLIFLKKDKRADV	506

The complete length ORF26ng nucleotide sequence <SEQ ID 695> is:

35	1	ATGAGCTGA TTGACTATT ACATTCATT TTCTCGGTG TGCCACCCCT	
	51	TTTGCACTG GCATCTGCC TCATTACCG CGCGCTACTG CTGCTTTAG	
	101	GCATCGTAT TTGGCTCGC GTTGCCTTT TGTCGGCGGC CAACCCCGTC	
	151	GACGGCTGA CACACTCGA AGACATGTC GTGCGCTTGC CTGGGAGA	
	201	CGGGATTG TGCGCTGGCA AACCAAAA CCTGGTTTTC CTGATACTTT	
	251	TGGCGATTTC CATTCTACTG CTGAGCTACT CGCGCAGAA TAAGCGTTT	
40	301	GCACGCTGGG CAAAAGCGCA CATTAAGCGA CGTGCCTGGCG CGAAAATGCT	
	351	GACCGCTCTG CTGCTGTTGC TAACCTTTA CGACGACTAT TTCCACAGCC	
	401	TGGCGCTTG TGCGATGCC GCCCCCGTTA CGGACAAGTT TAAAGTTTCC	
	451	CGCCCAAACAC TGCGCTCAT CCTCGACTCC ACTGCGCTGC CCATGTGCGT	
	501	GCTGATGCCG GTTCAAGCTG GGCGCCGGC GAITATGCCG AGCGTTGCCG	
45	551	GATTGCTCTG TACCTACCAA ATTACCGAA ACACGCCGAT GGGGACCTT	
	601	GCGCATGTA CGCTGATGAA CTATATCGG CTGTTTGGCC TGATTATGGT	
	651	ATTGCGTGTG GCATGTTCT CTCCTGACAT CGCGCTGAtc qCGCGTTGCG	
	701	AACAGGCTGC GTTGAACGGG gcccaaggqaa aaaaaaggccqcg tTCAGAcgCT	
50	751	ACCAAAAGTCG GTCTTTAGCG ATTEGATTATT CGCGTTTGG CCTTAATCGC	
	801	TCACAAACGGT TCCGGCATGA TCTACACCGG CGGGCAGGCC AGCGGAAACCT	
	851	TGACCATTTT GGGGGCACTT GAAAATACCC ACCTAAACAC TTCCCTGGTA	
	901	TTCCGGCGGA CTGGGGGGT GCTTGGCGTC GTCTCTGCA CTTCCTGCAC	
	951	GTTAAACACC GCGATGATATC CAAAAGCCGT GTGCGAGGGT GGAAATTCGA	
55	1001	TGTCGGCG AATGCCATT TTAAATCCCG CCTGGCTCAT CASTACCGTT	
	1051	GTCCGGCAAAT TGCACACGGG CGACTACCTC TCCACGGCTGG TTGGCGGCCAA	
	1101	CATCCCATCCC GGCTCTCTGC CGCGTACATC CTTCCTGCTC GCGACGCTGA	
	1151	TGGCGTTTGC CACAGCGCA AGCTGGGGGA CGTTCGGCGAT TATGCTGCCG	
	1201	ATTGCGCCGG CACATGGCGT CAAAGTGGAA CCCGGCGTGA TTAatccGTG	
60	1251	TATGCGCGA GTATAGGGC GGGGGTAGT CGGGCGGACAC TGTTCGCCCCA	
	1301	TCTCCGACAC GACCATCCCTG TGCTCACCG GCGCGCGCTG CACACCATC	
	1351	GACACGCTTA CCTCGAACCT GCCTTATGCC CTGACGGTTG CGGCCGCCGC	
	1401	CGCATCGGGC TACCTCGCAT TGGGTCTGAC AAAATCCGG CGTGTGGCT	
	1451	TTGGCACGAC CGGTATGTA TTGGCGTGC TGATTTTCT GTTGAAGAT	
	1501	AAAAAACGGC CGCACCTTG A	

65 This encodes a protein having amino acid sequence <SEQ ID 696>:

1 MQLIDYSHSF FSVVPPFLAL ALAVITRRLV LSLGIGILVG VAFLVGGNPV
 51 DGLTHLKDMDV VGLAWADGGW SLGKPKLILF LLLGIETSL LTYSGSNQAF
 101 ADWAKRHINR RCGAKMLTAC LVEFTFIDDY FHSILAVGAIA RPVTDKFKVS
 151 RAKLAYILDS TASENCVILMP VSSWGASTIA TLAGLILTYK ITEYTPMGTF
 201 VAMSIMNYYA LPALIMUVV AWFSFDIGSM ARFEQARALNE AQDETAASDA
 251 TKGRVIALII PVLAIIASIV SAMIYTGAQA SETFSIILGF AFNTDVNTSLV
 301 FGGTCGVLLAV VLCTFGTIRT ADYPAKAVWQG AKSNGFAIAI LILAWLISTV
 351 VGEHTYGDYL STLVAGNIHP GELPVILLELL ASVMFAFATGT SWTGFGLMP
 401 IAAAMAVKRVV PALIIPCMSSA VMAGAVCGDH CSPISDTIL SSTGARNCHI
 451 DHVTSQLPYA LTVAARAAASG YLALGLTKSA LLGFGTTGIV LAVLIFLLKD
 501 KKRADV*

ORF26ng and ORF26-1 show 98.4% identity in 505 aa overlap:

		10	20	30	40	50	60
15	orf26-1.pep	MQLIDYSHSFSSVVPFLAL ALAVITRRLV LSLGIGILVG VAFLVGGNPV DGLTHLKDMDV					
	orf26ng	MQLIDYSHSFSSVVPFLAL ALAVITRRLV LSLGIGILVG VAFLVGGNPV DGLTHLKDMDV	10	20	30	40	50
			70	80	90	100	110
20	orf26-1.pep	VGLAWSDGDSMSLGKPKLILF LLLGIFTSLLTYSGSNQAFADWAKRHINR RCGAKMLTAC					120
	orf26ng	VGLAWADGDSMSLGKPKLILF LLLGIFTSLLTYSGSNQAFADWAKRHINR RCGAKMLTAC	70	80	90	100	110
			130	140	150	160	170
25	orf26-1.pep	LVFVTFIDYDFHSLAVGAIA RPVTDKFKVSRKLAYILDSTAAPMCVLMVSSWGASIIIA					180
	orf26ng	LVFVTFIDYDFHSLAVGAIA RPVTDKFKVSRKLAYILDSTASP MCVLMVSSWGASIIIA	130	140	150	160	170
			190	200	210	220	230
30	orf26-1.pep	TLAGLLVTVYKITEYTPMGTFVAMSLMNYALFALIMFVUVVWAFSEFDIGSMAREPQARALNE					240
	orf26ng	TLAGLLVTVYKITEYTPMGTFVAMSLMNYALFALIMFVUVVWAFSEFDIGSMAREPQARALNE	190	200	210	220	230
			250	260	270	280	290
35	orf26-1.pep	AHDETAASDATKGGRVYALIIIPVLALIASIVSAMIYTGAQASETFSIILGAFENTDVNTSLV					300
	orf26ng	AHDETAASDATKGGRVYALIIIPVLALIASIVSAMIYTGAQASETFSIILGAFENTDVNTSLV	250	260	270	280	290
			310	320	330	340	350
40	orf26-1.pep	FGGTGCGVLAVVLCGLTGTIKTADYPKAWVQGAKSMFGAI AILILAWLISTVVGEMHTGDYL					360
	orf26ng	FGGTGCGVLAVVLCGLTGTIKTADYPKAWVQGAKSMFGAI AILILAWLISTVVGEMHTGDYL	310	320	330	340	350
			370	380	390	400	410
45	orf26-1.pep	STLVAGNTHPGFPLVILELLASVMAFATGTSGTGFIMLP IAAAMAVKVEPALIIPCMSSA					420
	orf26ng	STLVAGNTHPGFPLVILELLASVMAFATGTSGTGFIMLP IAAAMAVKVEPALIIPCMSSA	370	380	390	400	410
			430	440	450	460	470
50	orf26-1.pep	VMAGAVCGDHCSPISDTTISLSTGARNCHIDHVTSQLPYALTVAARAAASGYLALGLTKSA					480
	orf26ng	VMAGAVCGDHCSPISDTTISLSTGARNCHIDHVTSQLPYALTVAARAAASGYLALGLTKSA	430	440	450	460	470
			490	500			
55	orf26-1.pep	LLGFGTTGIVLVLALIPLLKKDKKKRANAX					
	orf26ng	LLGFGTTGIVLVLALIPLLKKDKKKRANAX	490	500			
60							
65							

In addition, ORF26ng shows significant homology to a hypothetical *H.influenzae* protein:

-400-

sp|P44263|YF86_HAEIN HYPOTHETICAL PROTEIN HI1586 >gi|1074850|pir||C64037
 hypothetical protein HI1586 = *Haemophilus influenzae* (strain Rd KW20) >gi|1574427|U32832| H.
 influenzae predicted coding region HI1586 [Haemophilus influenzae] Length = 519
 Score = 538 bits (1370), Expect = e-152
 Identities = 280/507 (55%), Positives = 346/507 (68%), Gaps = 7/507 (1%)

Query: 1 MQLIDYSHSFFSVVPPFLALALAVITRXXXXXXAFLVGGNPVDGLTHLKDMLW 60
 M+LID+S +S+V+ LA+ LA+ TRR L +L V
 Sbjct: 14 NELIDFSSSWWSIVPALLAIIIAIATRRLVLSLSAGIIIGS1M1SDWQIGSAFNLYLVKVN 73

Query: 61 VGLAWAIDGDWSLGKPKILVPLILLGIFTSTLTYGSQNQAFADWAIRHKRNRCGAIKMLTAC 120
 V L +ADG+ + I+FL+LLG+ T+LL+ SGSN+AFA+WA+ II R GAK+L A
 Sbjct: 74 VSLVLYADGEIN-SNNNIVLFLLLLGVLTALLTVSGSNRAFARWAQSRIKGRRGAKLIAAS 132

Query: 121 LVVFVTFIDDYFHSLAVGAIARPDTKFKVSRAKLAYILDSTASPMPCVLMPSWSGASIIA 180
 QVTFIDDYFHSLAVGAIARPVTD+FKVSKRALKAYILDSTA+PMCV+MPVWSWG A II
 Sbjct: 133 LVVFVTFIDDYFHSLAVGAIARPVTDRFKVSRAKLAYILDSTA+PMCVMMPSWSGAYIIT 192

Query: 181 TLAGLILWVITYKEITPYGTTGTVANSLNNYVALLIMVFFVVAWFSDIGSMARFEQAAALNE 240
 + GLL TY ITETPY+C EVAMS NN+YX+A+IMVF VA+FSEDI SM R E+ AL
 Sbjct: 193 LIGGLLALATSYTETPYGIAEVAMSMNMFYIAFISIIMVFFVAFSEFDIASMVHKEKLKLN 252

Query: 241 AQDETAASDATKGRVYALIIPVULIALIASTVSAMIYTGAQA---SETFSILGAFENTDD 296
 +D+ TRG+V LI+P+L LI +TWS MIYTGA+A + FS+LG FENT V
 Sbjct: 253 TEDQLEEETGTGKVQRNLILPIIIVLIIATWSMNIYTGAEALAADGKVFSVLGFENTVVG 312

Query: 297 TSLVFGGTCGVL--AVWLCTFGTIKTADYPKAVWQAKSMFGXXXXXXXXXXSTVVGEN 354
 TSLV GG C ++ +++ + + Y ++ G KSM G + +VG+N
 Sbjct: 313 TSLVGGCISIIISTLILIDRQVSVPEYVRSWIVG1KSNSGAIAILFFAWTINKIVNG 372

Query: 355 HTGYLSTLVAGNIIIPGFLFVILFLASVNAFATGTSWGTGFIGMLPIAAAMAVKVEPALI 414
 TG YLS+LV+GNI FLFVLPFL + NAF+TGTSGTGFICMLPIAAAMMA P L+
 Sbjct: 373 TQGKYLSSLVGNIPMQFLPVILEVGLGAAMAFSTGTSWGTGFIGMLPIAAAMAAAEP 432

Query: 415 IPCMSAVNAGAVCGDHCPSPISDTTILSSTGARNCHIDHVTQXXXXXX 474
 +PC+SAVMAGAVCGDHCPSP+SDTTILSSTGQ+CNHIDHVT+Q
 Sbjct: 433 LPCLCSAVNAGAVCGDHCPSPDITLSSSTGAKCNHIDHVTQLPYATVATATSIGYIVV 492

Query: 475 XXXKSALLGFTTGIVLAVLIFPLKDK 501
 S L GF T + L V+IF +K +
 Sbjct: 493 GFTYSGLAGFAATAVSLIVIIFAVVKR 519

Based on this analysis, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*,
 and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 83

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 697>:

1 ..AGAACATGGT ATGCCGACGN .AGTATCAAG ACGGAAATGG TTATGGTCAA
 51 CGATGAGCTT GCAAATTC TGACTTGGG TGAAAGCGGC CGATTACTT
 101 CGGAACGTC TATCCCCAC CATCAACGCA ACGGGGTGGT TTTGGAGTGG
 151 TATGAGATG GTTCTAAAAA GAGCGAAGT. GTTTATCAGG ATGACAAGTT
 201 GGTCAAGAAA ACCCAAGTGGG ATAAGGATGG TTATTTAAC GAACTCTGA

This corresponds to the amino acid sequence <SEQ ID 698; ORF27>:

1 ..KOWYADXSIIK TEMVWNDEP AKILTWDSEG RLLSELSIRH HQRNGVVLEW
 55 51 YEDGSKKXSE VYQDDKLVRK TQWDGDGYLI EP*

Further work revealed the complete nucleotide sequence <SEQ ID 699>:

1 ATGAAAGAAAT TATCTCGGAT TGTATTTCA ACTGTCTGT TGGGTTTC
 51 GCGCCCTTTC CGCGCGCAGA CCTATTCTGT TTATTTAAT CAGAACGGAA
 101 AGCTGACGGC GACGATGCT TCTGCCGCTT ATATCAGGCA ATATAGTGTG
 151 GTGGCGGGTA TTGCGACGCC GCAGGATT TATTATCGGT CGATGAAGAA

-401-

201	ATATTTCTGAA	CCTTATATCG	TTCCTTCAAC	GCAAAATCAA	TCTTTGTG
251	CTACCTCCGA	AAACGGTATG	TTGATTTTGT	GCGCATTTAA	TGGCCAGAAA
301	AAAATGGGGG	GGGGCCTCGA	CAAGGGTAAG	CGGGAGGGG	AGGGTGCCTG
351	CTGGTATCCA	ACAGCTTAAA	AACTCTGCCG	ATATGCTTAT	AAAAATGGCT
401	TAGTGGAGGG	TACGGGATAC	CGCTATTTAC	GTAAACGGCG	CAAGGAAAGC
451	GAATACTTCACT	TAAAGCAAC	TAAGGCCAAC	CGCGCTATGGA	AAGCAATGTC
501	TGGCCGCCCC	AGTATCAAGA	CGGAAATGTG	ATATGGCTCA	GATGGAGCTG
551	CCAAATTTCA	GACTTGGTG	GAAAGCCGCC	GACTACTCTC	GAAGCTCTG
601	ATACGGCCAC	ATCAACCGAA	CGGGGGTTGT	TTGGAGTTGT	ATGAAGATGG
651	TTCTAAGGAG	AGCGCACTGG	TTTATCTAGGA	TGACAAGTGTG	GTCAAGGAAA
701	CCCGAGTGGGA	TAAGGGATGT	TTATTAATCC	ACCCCTGGA	

This corresponds to the amino acid sequence <SEQ ID 700; ORF27-1>;

1 MKKLRSIVFS TVLGGFSAAL PAQTYSVYFN QNGKLTTATMS SRAVYIROS
 51 VAGIAHQAQD YYPSMVKYSE PIYAVSTCKE SFVPTFLQNMG LILWHNGQVS
 101 KMGAFGSKKE PDGEWNVWYNG NGKSLSAVMV KNGLSGTGTV YLWHNGKKS
 151 EIQFFQKAN GWQWAKDAG SIKTEMVNMN DEPAKLTWD ESGRLLSELS
 201 IRHORHNGVV LEWYEDGSKK SEAVQDKDL VRKTQWDKD LYIIEP

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

20 ORF27 shows 91.5% identity over a 82aa overlap with an ORF (ORF27a) from strain A of *N. meningitidis*:

The complete length ORF27a nucleotide sequence <SEQ ID 701> is:

35	1	ATGAAATAAT	TATCTGGT	TGTATTTCGA	ACTGCTCTGT	TGGTTTTTGT
	51	GGGGCGCTTA	GCGGCAGCA	NCTATTCTGT	TTATTTAT	CAGAACGGGA
	101	AACTGACGCC	GACGNTGCT	TCTGCGCNC	ATATCAGGCA	ATATAGTGTG
	151	GGCGGAGGTA	TTCGGCACCS	GCAGGANTT	TANTATCTCGT	CGAGAAGAAN
	201	ATATCCCGAA	CTTATATCG	TTCGTCIAAC	GCACAAATCA	TCTTTTGCGC
40	251	CTACCTCCGA	AAACGCTTG	TGTGATTGTTG	GGCATTTTAA	NGCTGAGAAC
	301	AAAATGGCNC	GGGGCTTCGA	CAAGGGTAAG	GGCGACGGG	AGTGGTGTGAC
	351	CTGTGTTAACG	AAACGGTAA	AATCTGCGGT	TATGCTTAT	AAAATGGTGT
	401	TGAGTCAGGC	TACCGGGTTN	CGTATTTAC	TAACCGCGA	CAAGGAAAGC
	451	GAATACTTCAGT	TTTAAACGAA	TAAGGCAACG	GGCTATGTGA	AAACGATGTA
	501	TGGCCGACGG	AAATATCAA	GGAAATGTGT	TATGTCGAT	GATGAGGCTTG
45	551	CCAAATACTT	GACATGGT	GAAGACGGTC	GATTACTCTC	GAAGCTGTCT
	601	ATTCATCATC	ATNAACCTA	TGGAGTAGTC	TTAGAGTGTG	ATGAAGATGG
	651	TTCTTAAAGG	ANTGAAGTC	TTTATCAGGA	TGATAAAGTGT	GT CAGGAAAA
	701	CCCAAGGGAA	TAANGATGT	TATTTATCAG	AAACCTGGA	

This encodes a protein having amino acid sequence <SEQ ID 702>.

```

50      1 MKKLLSIRIVFS TVLLGFSAA PAQKYSVYFN QNGKLATKS SAAVIRQYSVW
51     51 AEGIAJGKX FYPMSKWWYE PIYVASTQIK SFVFTPLQNM LILWIFHGKQ
100    101 KMGAGFSKKGK PGGEMWNWTP NGKGSAPMVY KNGLSEGTGX RYNNWFGKQ
151   151 EIQFKONQAN GWKWWQDAGK NIKTEMVNWS DEPAKILWTG ERSGLLS
200   201 THHWWKGVW LEWEDGSKK XRAVYODDKI VRTKTDWMDG YLTJRP

```

55 ORF27a and ORF27-1 show 94.7% identity in 245 aa overlap:

orf27a.pep MKKLSRIVFSTVLLGFSAAALPAQXSYVFNONGKLATXXSAAYIQRQSYVAEGIAHAQXE

	orf27-1	MKKLSRIVFSTVLLGFSALPAQTYSVYFNQNQKLATMSSAAYIRQYSVVAQIAHAQDFE
10		10 20 30 40 50 60
5	orf27a.pep	XYPSMRKVKSEPYIIVASTQIKSFVPTLQNQMLILWIFKGQKRMAGGFSKGPDGEVNWYP
	orf27-1	YPPSMRKVKSEPYIIVASTQIKSFVPTLQNQMLILWIFKGQKRMAGGFSKGPDGEVNWYP
10		70 80 90 100 110 120
orf27a.pep		NGKKSADVMPYKNGLSEGTTGKRYYYRNGKESEIQFKQNKAANGVWKQWYADGSIKTEMVVMN
15	orf27-1	NGKKSADVMPYKNGLSEGTTGKRYYYRNGKESEIQFKQNKAANGVWKQWYADGSIKTEMVVMN
	orf27a.pep	130 140 150 160 170 180
	DEPAKILTWDESGRLLSELSIIHHHKRNQGVLEWYEDGSKKEAVYQDDKLVRKTQWDKDG	140 150 160 170 180
20	orf27-1	DEPAKILTWDESGRLLSELSIIHHHKRNQGVLEWYEDGSKKEAVYQDDKLVRKTQWDKDG
		190 200 210 220 230 240
25	orf27a.pep	YLIEPK
	orf27-1	YLIEPK

Homology with a predicted ORF from *N.gonorrhoeae*

ORF27 shows 96.3% identity over 82 aa overlap with a predicted ORF (ORF27ng) from

30 *N.gonorrhoeae*:

	orf27.pep	KQWYADXSIKTEMVVMNDEPAKILTWDESG	30
	orf27ng	LSEGTTGKRYYYRNGKESEIQFKQNKAANGVWKQWYADGSIKTEMVVMNDEPAKILTWDESG	193
35	orf27.pep	RLLSELSELRIHHKRNGNVLEWYEDGSKKEAVYQDDKLVRKTQWDKDGYLIEP	82
	orf27ng	RLLSELSELRIHHKRNGNVLEWYEDGSKKEAVYQDDKLVRKTQWDKDGYLIEP	245

The complete length ORF27ng nucleotide sequence <SEQ ID 703> is:

40	1 ATGAAGAAAT TATCTCCGAT TGTATTITCA ATCGTACTGT TGGGTTTTTC 51 GGGCGCTTTC CGGGCGCAGA CCTATTCCTGT TTATTTTAAT CAAACACGGG
	101 AACTGACGGG GACGATGCTT CTGCGCGCTT ATATCAGCGG ATATAGTGTG 151 GCGGCGCGTA TGCGCACACG CGACGATTTT TATTATCCGT CGATGAGAAA
	201 ATATCTCGGA CCTTATATCG TTGCTTCAAC GCAAATCAA TCCTTTGTC 251 CTACCTCTCA AAACCGTAGT TTGATTTTGT GCATTTTAA TGGTCAGAAA
	301 AAAATGGCGG GGGGCTTCGA CAAGGTTAAC CGCGACCGGG AATGGCTCAA 351 CTGCTATCGC AACGGTAAAAA AATCTGCCGT TATGCCCTAT AAAAATGGCT
45	401 TGAGTGAGGG TACGGGATCT CGTTTATACG GTAACGGCGG CAAGGAAAGC 451 GAAATCCAGT TTAAAGCAAGA GGACGGAC ACAGGATGGAA AGCAATGGTA
	501 TGCCGATGGA AGTATCAAAA CGGAAATGGT TATGGTCAAC GATGAGGCTG 551 CCAAAATCTC GACTGGGAT GACTACTTC GGAACTGCT 601 ATCCGCACCA ATAACACCAA CGGGGTGGTT TTGGAGTGGT ATGAAGATGG 651 TTCTAAAAAG AGCGAGGCTG TTATACAGG TGACAAGTTG GTCAAGAAA
50	701 CCCAATGGGA TAAGGATGGT TATTAATACG AACCTGTA

This encodes a protein having amino acid sequence <SEQ ID 704>:

55	1 MKKLRSRIVFSTVLLGFSALPAQTYSVYFNQNQKLATMSSAAYIRQYSVVAQIAHAQDF 51 AAGIAHAQDF YXPMSMKYSE PYIIVASTQIK SFVPTLQNMG LILWIFHNGOK
	101 KMAFGESKKG PDGEVNWYP NGKKSADVMPY KNGLSEGTTG RYVYRNGKES 151 EIQFKDNKAN GWKMKWYADG SIKTEMVMN DEPAKILTD ESGRLLSELS
	201 IRHHKRNGVV LEWYEDGSKK SEAVYQDDKL VRKTQWDKDG* YLIEP*

60 ORF27ng and ORF27-1 show 98.8% identity in 245 aa overlap:

	orf27-1.pep	10 20 30 40 50 60
	MMKLSRIVFSTVLLGFSALPAQTYSVYFNQNQKLATMSSAAYIRQYSVVAQIAHAQDF	

	orf27ng	: : : : : : : : : : MKKLSRIVFSIVLLGFSAAALPAQTYSVYFNQNQKLTAATMSSAAYIRQYSVAAGIAHQDF 10 20 30 40 50 60
5	orf27-1.pep	70 80 90 100 110 120 YYPSMVKYSEPYIVASTQIKSFVPTLQNQMLILNHFNGQKVMAGGFSKGPDPGEVNWYP
	orf27ng	: : : : : : : : YYPSMVKYSEPYIVASTQIKSFVPTLQNQMLILNHFNGQKVMAGGFSKGPDPGEVNWYP 70 80 90 100 110 120
10	orf27-1.pep	130 140 150 160 170 180 NGKKSAVMPYKNGLSEGTGYRYYRNNGKESEIQFKQNKAANGVWKQWYADGSIKTEMVWN
15	orf27ng	: : : : : NGKKSAVMPYKNGLSEGTGYRYYRNNGKESEIQFKQNKAANGVWKQWYADGSIKTEMVWN 130 140 150 160 170 180
	orf27-1.pep	190 200 210 220 230 240 DEPAKILTWDESGRLLSSELISRHHQRNGVLEWYEDGSKSEAVYQDDKLVRKTQWDKD
20	orf27ng	: : : : DEPAKILTWDESGRLLSSELISRHHQRNGVLEWYEDGSKSEAVYQDDKLVRKTQWDKD 190 200 210 220 230 240
25	orf27-1.pep	YLIEPX
	orf27ng	YLIEPX

Based on this analysis, including the putative leader sequence in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF27-1 (24.5kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 17A shows the results of affinity purification of the GST-fusion protein, and Figure 17B shows the results of expression of the His-fusion in *E.coli*. Purified GST-fusion protein was used to immunise mice, whose sera were used for ELISA, which gave a positive result, confirming that ORF27-1 is a surface-exposed protein and a useful immunogen.

Example 84

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 705>:

40	1	ATGAAATTAA CCAAGCACCC CGTCTGGGCA ATGGCGTTCC GCGCATTTTA 51 TTTCGCTGGG GCTCTGACG CGCGATTTGC CGTATTTGCTG TGGGGTTTCG
	101	GCTACAOGGG AACGCAACAG CTGTCGCGTT TCTATTTGCA CGCCATGAG
	151	ATGATTTGGG GTTATGCCGG ACTGTCGTCG ATGCGCTTCC TGTGACGCC
	201	CGTGGCCACT TGAGCGGGGC AGCGGCCCCAC CGGGGGGGGC GtaTCGTGTC
	251	GCGCTTACTA TCTTTTGCGT GGCTGGGGGG ATTGCGCGCT TTATCCCGGG
	301	TTGGGGTGGC TCAGCAAGCG GCATACTCGG TACGCTGTT TTCTGGTAGG
	351	GGGGGGTGT CATGGCTTCCC CCCCTTATCC GTTCCGAGAA TCAACCGAAC
	401	TATGTTgCCG TGTTGGCGCT GTTGTCTTGC CGGGCACCG ATGGGGCGT
	451	CCAGCTTCCAG CTGCAACACG GCAACCTAGG CGGACTCTTG AGCGGATTGC
	501	AGTCGGGCTT GGTGATG

50 This corresponds to the amino acid sequence <SEQ ID 706; ORF47>:

1	MEKFTKHPVWA MAERPFYSLA ALYGALSVLL WGFGYTGTGX LSGFYWHAE
51	MIWGYIAGLVV IAFLLTAVAT WTGPQPFTRGG VLVLGLTIFWL AARIAAFIPG
101	WGASASGILG TLFWYGAWC MALPVIRSQN QRNYWAVFAL FVLGGTHAFA

151 HVQLHNGNLG GLLSGLQSQL VM

Further work revealed the complete nucleotide sequence <SEQ ID 707>:

5 1 ATGAAATTTA CCAAGCACCC CGTCTGGGA ATGGCGTTCC GCCCATTTA
 51 TTGCGTGGCG GCTCTGTACG GCGCATGTC CTATTTGCTG TGGGGTTTCG
 100 101 GCTAACCGG AACCACAGG CTGTCGGTT TCTATTGCGA CGGCATGAG
 151 151 ATGATTCTGG GTTATGCCGG ACTGGCTGTC ATCCGCTTCC TGTGACCCGC
 201 201 CGTCGCACAT TGAGCGGGC AGCGCCACAC GGGGGGGGGC TTCTGTCG
 251 251 GCTTGACTAT CTTTGGCTG GTGCGCGCGA TTGGCGGCTT TATCCCGGGT
 301 301 TGGGGTGGCT CGCGAACGGG CAACTCTGGT ACCTGTITTT TCCTGGTACGG
 351 351 CGCGGTGTGG ATGGCTTGC CGCTTATTCGG TTCCGAGAAT CAACGCAACT
 401 401 ATGTCGGCTG TGCGCCCTG TTGCTCTGG CGGCACGCA TGCGCGGTC
 451 451 CACGTCGAGC TGCAACACGG CAACCTAGGC GGACTCTTGAA GCAGGATGCA
 501 501 GTCGGGCTG GTGATGTGT CGGGTTTATAC CGGCTCTGATT GTGACCGGGA
 551 551 TTATTTCTTCTT TTATGCTCC AAACGCTTGTG ATGTCGGCGA GATTCGGAGT
 601 601 CGGAAATGGG TGGCGGAGG TTCCGCTGGG CTGCCATGCG TTGACTGCCAT
 651 651 GCTGTGATGCC CACGGTGTGT TGGCTTGGCT GTCTGCCGTG TTGCGCTTTC
 701 701 CGGGCAGTGT GATTTTACG GTCAGGTGT ACCCTGTGIG GTATAAACCC
 751 751 GTGTTGAAAG AGCCGATGGT GGGATCTG TTGGCGGCTT ATCTGTTAAC
 801 801 CGGGATGGGG CTGATGCGG TCAGCGGCTC TTATTCGAAA CGCGCTTTC
 851 851 TCACTCTGGG TGTGCTATCG ATCGGGGTCG CGGGTATCGG CGTGTGACT
 901 901 TTGGGGATGA TGCGGCTCAT CGCGCTTGGT CATACGGGCA ATCCGATTTA
 951 951 TCGCGGGCC AAAGCGCTTC CGCTTGCCTT TTGGCTGATG ATGGCGGCAA
 1001 1001 CCGGCCCTGG TAGTGTGCG TTGTTTTCCTT CGGGCAGTGC CTACACGGC
 1051 1051 AGCATCCCGA CCTCTCTGGT TTGTTTGCAT CGGGCGTTT TTGTTGATG
 1101 1101 GTGGAAAGTAT ATTCCTTGGC TGATTGCTG CGGTTGCGGAC GGCAAGCCCG
 1151 1151 GTTGA

This corresponds to the amino acid sequence <SEQ ID 708; ORF47-1>:

30 1 MKFTKHPWVA MAFRPFYSLA ALYGALSVL WGFGYTGTHE LSGFYWHAAE
 51 MINGYAGLVV IAFLLTAVAT WTGQPPTTRGGV VLVLGTLIFWL MARIAAFIPG
 101 101 WGASASGLL TLFFWYGAVC MALPVIRSNQN QRNIVAVFAFL FVLGGTHAAF
 151 151 HVQLHNGNLG GLLSGLQSQL VMVSGFIGLI GTRIISEFTS KRLNVPQIIPS
 201 201 PKWKAQASLW LPMLTAMLMA HGVLAWSAV FAFAAVGVIET VQVYRWYIKP
 251 251 VLKEPMWLIL FAGLYFTGGL LIAVGASYFK PAFLNLVHL IGVGGIGVLT
 301 301 LGMARTALG HTGNPIYPPF KAVPVAFLM MAATAVVRMVVA FVSSGTYATH
 351 351 SIRTSSVLF A LALLVYAWKY IPWLIRPRSD GRGP*

Computer analysis of this amino acid sequence predicts a leader peptide and also gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)ORF47 shows 99.4% identity over a 172aa overlap with an ORF (ORF47a) from strain A of *N.*

40 meningitidis:

		10	20	30	40	50	60
	orf47.pep	MKFTKHPWVA MAFRPFYSLA ALYGALSVL WGFGYTGTHE LSGFYWHAAE					
	orf47a	MKFTKHPWVA MAFRPFYSLA ALYGALSVL WGFGYTGTHE LSGFYWHAAE					
45	orf47a	MKFTKHPWVA MAFRPFYSLA ALYGALSVL WGFGYTGTHE LSGFYWHAAE	10	20	30	40	50
	orf47.pep	TAFLLTAVATWTGQPPTTRGGV VLVLGTLIFWL MARIAAFIPG	70	80	90	100	110
	orf47a	TAFLLTAVATWTGQPPTTRGGV VLVLGTLIFWL MARIAAFIPG	70	80	90	100	110
50	orf47a	TAFLLTAVATWTGQPPTTRGGV VLVLGTLIFWL MARIAAFIPG	70	80	90	100	110
	orf47.pep	MALPVIRSNQN RYVAFLVFLGGTHAAFHVQLHNGNLG LGLSGLQSQLVM	130	140	150	160	170
	orf47a	MALPVIRSNQN RYVAFLVFLGGTHAAFHVQLHNGNLG LGLSGLQSQLVM	130	140	150	160	170
55	orf47a	GRTRIISEFTSKR1NVPQI PSPKWAQASLWLPMLTAMLMA HGVMPWLSAFAFAAGVIFT	190	200	210	220	230
60	orf47a	GRTRIISEFTSKR1NVPQI PSPKWAQASLWLPMLTAMLMA HGVMPWLSAFAFAAGVIFT	190	200	210	220	240

The complete length ORF47a nucleotide sequence <SEQ ID 709> is:

1 ATGAAAGCC CCAAGCAGG CGTGGGGCA ATGGCGTCG CCGCTTTTA
 5 TTACTGGGC GCTCTTAGC GGCATCTTG CTGATTGTCG TGCGGTTCG
 101 GCTACAGGG AACGCACGAG CTGCGCGGT TCTATTGCGA CGCCGATGAG
 151 ATGATTTGGG GTTATGCCG ACTGTCGTC ATGCGCTTC TGTCGACC
 201 CGTCGCACT TGGACGGGC AGCGCCCGC CGGCCCCGGC GTTCGTGCG
 251 GTGTCAGAT TTTGGCGGCT GTGCGGGGA TTECCGCTTC ATTCGGCG
 301 TGGGGCTC CGGCAAGCGG CTACTCGGT AGCTGGTTC TTCTGGTACG
 351 CGCGCTTCG ATGGCTTGG CGGTATTCGG TTCCGAGAT CGCGCTTAC
 401 ATGTTGGCT GTTGGGGCT TTCTGCTGG GGGTACCCCA CGCGCTTAC
 451 CAGCTCCAGG TGCACACGG CAACCTAGGC GGACTCTTGG GAAGCTTAC
 501 GTCGGGCTC GTGATGGTGT CGGGTTTAT CGCTGTGATG GTACGGGA
 551 TTTATGGTGT TTTATGGC AACGGTGTG ATGTCGGCA GTTCCGAT
 601 CGCAATGGG TGGCGGAGG TTGCGTGTG TTCCGCTTC TGACCGCTAT
 651 GTGTCAGGG CACCGCGTC TTGCGTGTG TTCCGCGCTC TTTCGATCT
 701 CGCGCAGGT GTATTITACG GTGCAAGCTG ACGCGTGGT GTAAAGCTT
 751 GTGTGAAAG AGCCGATGCT GTGGATCTG TTTCGGCT ATCTGTTAC
 801 CGGATGGGG CTGATGGTG CGGGCGGCT TTATTCGAA CCCTGGCTTC
 851 TAATCTGGG TTGTCATCTC ATCGGGCTCG CGGTGATTCG CTGGTACAT
 901 TTGGCGATG TGCGGGCTC CGCCGTCGGT CATACGGCGA CGCATTTTA
 951 TCGCCGCCA AAAGCGCTG CGCTTGGCT TTGGCTGATG ATGGCGGCA
 1001 CGCGCGTCCG TATGTTGGT CTGGCTTCGCGT CCGGCACTGC TCAACCGC
 1051 AGCATACGA CCTCTGGGT TTGTTGTCG CCGGCGCTTC TTGGTATGEC
 1101 GTGGAGAT ATTCTTGGC TGATTCCTCC CGCTTGGAC GGCAAGCCCG
 1151 CTGAA

This encodes a protein having amino acid sequence <SEO ID 710>;

1 MKPFPKFWA MAFPPPFYLLA ALYGVSLWL WGGFTGHTLNS LSGFVWHAIE
 5 MINCYGLVU TLFLTTAVAT WTQCPPTRCG VVGLTILWFL AARIAAFIGP
 101 WGACASAGLG TLFWXWGGCA MALPVIRSON ORNVAEYAWF VLFGVTHAA
 151 HVOLLINGENLGL GLLSLQLSGL VVMSCFGIHL CIRLISIFTS KRLNVPOIPS
 201 PKWAQASLM LPMLTATMLM HGVMWPLSRA PFAAFAGVQ YQVRYWYKQ
 251 VKEPLMWLFL FAGYLPTGLG LIAYAGSYFK PFLANLGVHL IKGVGICVL
 301 LGMMARTLG HIGNPYLIPP KAVPUFWLM MAATAVRMVLS VFSSGTAYTH
 351 SIRTSYSLFLA IALLVYVPLC WMLPRLRSTL GRFG*
 30

35 ORF47a and ORF47-1 show 99.2% identity in 384 aa overlap:

		10	20	30	40	50	60
40	orf47a.pep	MKFTKHPVWAMAFRPFYSLAALYGALSVLWLWGPGYTGTGTHLSGFYWHAEHSMIWGYAGLVV					
	orf47-1	MKFTKHPVWAMAFRPFYSLAALYGALSVLWLWGPGYTGTGTHLSGFYWHAEHSMIWGYAGLVV	10	20	30	40	50
45	orf47a.pep	IAFLLTAVATWTGQPPTTRGVLYVGLTIFWLAARIAAFIPGWCASASGLCTLFFWYGAVC	70	80	90	100	110
	orf47-1	IAFLLTAVATWTGQPPTTRGVLYVGLTIFWLAARIAAFIPGWCASASGLCTLFFWYGAVC	70	80	90	100	110
50	orf47a.pep	130 140 150 160 170 180					
	orf47-1	MALPVIIRSNQNQNYVAVFALFVLLGGTHAAHFHVQLHNGNLGLGSQGLVMVSGFIGLI					
55	orf47a.pep	MALPVIIRSNQNQNYVAVFALFVLLGGTHAAHFHVQLHNGNLGLGSQGLVMVSGFIGLI	130	140	150	160	170
	orf47-1	GTRIIISFTTSKRNLNPQIPSPKWVAQASLWLPMLTAMLMAHGVPMWLSAFAFRAVGIVF	190	200	210	220	230
60	orf47a.pep	GTRIIISFTTSKRNLNPQIPSPKWVAQASLWLPMLTAMLMAHGVPMWLSAFAFRAVGIVF	190	200	210	220	230
	orf47-1	VQVYRWWYKPVLPKEPMLWLFAGYLTFGLGLIAVGASYFKPAFLNVLGWHLIGVGGIGVLT	250	260	270	280	290
65	orf47a.pep	VQVYRWWYKPVLPKEPMLWLFAGYLTFGLGLIAVGASYFKPAFLNVLGWHLIGVGGIGVLT	250	260	270	280	290
	orf47-1	VQVYRWWYKPVLPKEPMLWLFAGYLTFGLGLIAVGASYFKPAFLNVLGWHLIGVGGIGVLT	310	320	330	340	350

-406-

	orf47a.pep	LGMMAARTALGHTGNPIYPPPCKAVPVAFWLMMMAATAVRMVAVFSSGTAYTHSIRTSVLP	
	orf47-1	LGMMAARTALGHTGNPIYPPPCKAVPVAFWLMMMAATAVRMVAVFSSGTAYTHSIRTSVLP	
5		310	320
	orf47a.pep	370	380
	orf47-1	IALLVYAWKRYIPWLIRPRSDGRPGX	
10		370	380
	orf47a.pep	IALLVYAWKRYIPWLIRPRSDGRPGX	

Homology with a predicted ORF from *N.gonorrhoeae*

ORF47 shows 97.1% identity over 172 aa overlap with a predicted ORF (ORF47ng) from *N.gonorrhoeae*:

15	ORF47	MKFTKHPVWAMAPRPPYSIA ALYGALSVLL WGFGYTGTGTHE LSGFYWHAE MIWGYAGLVV	60
	ORF47ng	MKFTKHPWAMAPRPPYSIA ALYGALSVLL WGFGYTGTGTHE LSGFYWHAE MIWGYAGLVV	60
20	ORF47	IAPLITIATAVWTWQGPPTRGGVVLVGLTIEWLALARIAAFIPFGWGAAASGLGILGTLPEFWYGAVC	120
	ORF47ng	IAPLITIATAVWTWQGPPTRGGVVLVGLTIEWLALARIAAFIPFGWGAAASGLGILGTLPEFWYGAVC	120
	ORF47	MALPVIRSNQRNRYVVAFALFVLGGTHAAFHVQLHNGNLGGLLSGLQSGLVM	172
25	ORF47ng	MALPVIRSNQRNRYVVAFALFVLGGTHAAFHVQLHNGNLGGLLSGLQSGLVMVWGFIGLI	180

The ORF47ng nucleotide sequence <SEQ ID 711> is predicted to encode a protein comprising amino acid sequence <SEQ ID 712>:

1	MKFTKHPVWA	MAFRPPYSIA	ALYGALSVLL	WGFGYTGTGTHE	LSGFYWHAE
51	MIWGYAGLVV	IAPLITIATAVWTWQGPPTRGGV	VLGQPTRGV	VLVGLTAFWML	AARIAAFIPFG
101	WGAAASGLG	TLFFWVGAVC	MALPVIRSON	RRNYVVAFAI	FVLFGGTHAA
151	HVQLNLNGL	GLSLQSQL	VMVWGFIGLI	GMKLIKSPFTS	KRLKLPOIPS
201	PKWVHASLW	LPMLNAlIMA	HRVMPWLSAA	PFPAAGVIFT	VQVYAGGITP
251	IEETSCGSVA	GICYRLGNSS	G		

The predicted leader peptide and transmembrane domains are identical (except for an Ile/Ala substitution at residue 87 and an Leu/Ile substitution at position 140) to sequences in the meningococcal protein (see also *Pseudomonas stutzeri* orf396, accession number e246540):

40	TM segments in ORF47ng				
	INTEGRAL	Likelihood = -5.63	Transmembrane	52 - 68	
	INTEGRAL	Likelihood = -3.88	Transmembrane	169 - 185	
	INTEGRAL	Likelihood = -3.08	Transmembrane	82 - 98	
	INTEGRAL	Likelihood = -1.91	Transmembrane	134 - 150	
	INTEGRAL	Likelihood = -1.44	Transmembrane	107 - 123	
	INTEGRAL	Likelihood = -1.38	Transmembrane	227 - 243	

Further work revealed the complete gonococcal DNA sequence <SEQ ID 713>:

45	1	ATGAAATTTC	CCAAACATCC	CCTGCTGGCA	ATGGCGTTCC	GCCCGTTTTA
	51	TTCACTGGC	GCATCTGAG	GGCCATCTGTC	CSTATTTGCG	TGGGGTTTCG
	101	GCTACACGG	AAACCGACGA	CTGTCGGGT	TCTATTTGGCA	CGGGCATGAG
	151	ATGATTTGGG	GTATTCGCG	TCTCTGCGTC	ATCCGCTTCC	TCTGTACGCG
	201	CGTCCCCACT	TGGACGGGGA	AGCCCGCCAC	GAGGGGCGGC	GTTCCTGGTCG
	251	GCTTGACGCC	CTTTGGCTG	GCTGGCGGGA	TTCGGCCCTT	TATCCCGGGT
	301	TGGGGTGGCG	CGGCAACGG	CATACTGGT	ACCTGTTTT	TCTGGTACGG
	351	CGCGGTGTGG	ATGGCTTTCG	CGCTTATCGC	TtcgCAAAAC	CGGCCCAACT
	401	ATGTCggCGGT	ATTCGCAATA	TTCGTCG	GGGTGCTCGA	TGGggcggtTC
	451	CACGTCggc	TGCAACACGG	AAACCTAGGC	GGACTCTTGA	GGCGATTTGCA
	501	GTGCGGCCGTG	GTATTCGCTG	CGGGCTTATT	CGGCGCTGAT	GGGATGAGGA
	551	TTATTTCGTT	TTTTCAGCTCC	AAACGGTTGA	ACGTGGCGCA	GATTCCCGAT
	601	CCGAAATGGG	TGGCGCAGGC	TTCGCTGCG	CTACCCATGC	TGACCCGCAT

-407-

```

651 AACTGATGGCG CACGGCGTGA TGCCTTGCGT GTGGCGGGCT TTGGCGTTTG
701 CGGGGGCGT GATTTCGAC GTACAGGTGT ACCGCTGGTG GTATAAACCC
751 GTATTGAAAG AACCCGCGCT GPGGATTCG TGTCGGCGCT ATCTGTTTAC
801 CGGATTGGGG CTGATTGCGG TTGCGCGCTC TTATTTCAAA CCTGCGTTCC
5 851 TCAATCTGGG CGTACACCTCG ATCGGGCTCG GCGGTATCGG CGTGCCTGACT
901 TTGGGCATGTA TGGCGCGTAC CGGGCTCGGT CATACCGGCC ATTGGATTTA
951 TCCGGCGGCC AAAGCGCTTC CGGTTGGTT TTGGCTGATG ATGGCGGCAA
1001 CGCCCGCTCGG TATGGTTGCC GTATTTCCT CGGGACTCG CTACAGGCC
1051 AGCATCGCA CGTCTCGGT TTGTTTGGAA CTGGCGCTGC TTGTTGATGC
1101 GTGGAATAAC ATTCCGTTGGC TGATCCGTCG GCGTTCCGGAC GGCAAGGCCG
10 1151 GTTGA

```

This encodes a protein having amino acid sequence <SEQ ID 714; ORF47ng-1>:

```

1 MKFTKHPWVA MAFRPPYSLA ALYGALSVL WGFGYTGTTHE LSGFYWHAAHE
51 MINWGYAGLVV IAIFLFLTRG WTGGPPTTRGG VLVGLTAWL AARIAAFIGP
15 WGAASAGILG TLEFWYNGAVC MALPVIRSON RBNYYAVFAI FVLGTHAAF
101 HVOIHLNGLG GLLSCLGOSL VMWSGFIGLII GMRIISFTTS KRLNVPQIPS
151 PKWVQAASLW LMFLTAIHLA HGWPMWLSAA FAFRAAGVIFT VQVYRWWYXP
201 VLFEPFLWIL GAFLYLTFCG LIAVGASYFK PAFLNLGVHL LGVGIGIVLT
251 LGMMARTALG HGNSTIIPPF KAVFVAFLWMA MARATAVRMVA VFSSGTAYTH
301 351 STRTSSSVLFA LALLYWAKY IPWLIRRSRD GRPG*

```

ORF47ng-1 and ORF47-1 show 97.4% identity in 384 aa overlap:

	10	20	30	40	50	60		
orf47-1.pep	MKF	KHPWVA	MAFRPPYSLA	ALY	GALSVL	WGFGYTGT	THE	LSGFYWHAAHE
25 orf47ng-1	MKF	KHPWVA	MAFRPPYSLA	ALY	GALSVL	VLLWGF	GTGTH	ELSGFYWHAAHE
	10	20	30	40	50	60		
orf47-1.pep	IAFL	LTTAVATWTG	QPPTTRGG	VGL	TLA	RIAA	IFP	GWNGASAS
30 orf47ng-1	IAFL	LTTAVATWTG	QPPTTRGG	VGL	TLA	RIAA	IFP	GWNGASAS
	70	80	90	100	110	120		
orf47-1.pep	MA	LPVIRSON	QNRNYVAV	FAL	FV	LG	GTG	TH
35 orf47ng-1	MA	LPVIRSON	QNRNYVAV	FAL	FV	LG	GTG	TH
	130	140	150	160	170	180		
orf47-1.pep	GTR	IISFFTS	SKRLNV	QPI	SPK	WVA	QASLW	LPM
40 orf47ng-1	GMR	IISFFTS	SKRLNV	QPI	SPK	WVA	QASLW	LPM
	190	200	210	220	230	240		
orf47-1.pep	VQV	YRWWY	KPV	LKE	PML	WIL	FEAGYL	FTG
45 orf47ng-1	VQV	YRWWY	KPV	LKE	PML	WIL	FEAGYL	FTG
	250	260	270	280	290	300		
orf47-1.pep	LGMM	MARTAL	GHITGN	PYI	PPP	KAV	VAF	WLM
50 orf47ng-1	LGMM	MARTAL	GHITGN	SITY	PPP	KAV	VAF	WLM
	310	320	330	340	350	360		
orf47-1.pep	370	380						
60 orf47ng-1	370	380						

Furthermore, ORF47ng-1 shows significant homology to an ORF from *Pseudomonas stutzeri*:

gnl|PID|e246540 (273914) ORF396 protein [Pseudomonas stutzeri] Length = 396
65 Score = 155 bits (389), Expect = 5e-37

-408-

Identities = 121/391 (30%), Positives = 169/391 (42%), Gaps = 21/391 (5%)

Query: 7 PWAMANFRPFYSLAALYGLSVALLNWGFYTGTGTHESGFY-----WHAHEMINTGYAGLV 59
 P+W +AFRPF+ +LY L++ LW +TG GF WH HEM++G+A +
 5 Sbjct: 14 PIWLAFLRFFFILAGSLYALLAIPLNVAATGLWP--GFQPTGGWLAWRHREMLFGPAMAI 71

Query: 60 VIAFLLTAVATWTGQPPTTRGGVLVGLTAFWLAARIAAFIPIPGWGAASGILGTLFFWYAV 119
 V FLLTAV TWTOQ G LVGA A WIAR+ ++ G AA L LF
 10 Sbjct: 72 VAGFLLTAVQTWTGQ?APSGNRLVGLAAWLARL-GWLFGLPAAWLAFLPLDLLFLVALVW 130

Query: 120 CNALPVIQRNQNRRNYVAVPAIFVLCGTHAAFXXXXXXVXXXXXVXXXXXMSVGFIQL 179
 MA + + +RN Y V + + G +V+ + L
 Sbjct: 131 MMAGLAWVRQKRNYFIVVLSLSMLGADVLILTGLLQGNDAALQRQGVLAGIWLVAALML 190

Query: 180 IGMRIISFFTSKRLNVPQIPSP-KWVAQASLWLPMITAIIAMHGCV---MPWLSAFAFA 234
 IG R+I FFT + L P W+ + A+L A GV PL F A
 Sbjct: 191 IGGRVVPLQRGLKVDAVKPWNWLDAVLLVTGVTGIVALLHAFGVAMRPQPLLFLV-A 249

Query: 235 AGVIFTQVYRWYKPVLPKEPMILWFLPAGYLFTGLGLIAVGASYE-KPAFXXXXXXXX 293
 GV +++ RW- K + C + LW L L+ + + + + F A
 Sbjct: 250 IGGVGHLLRMRWYDKGIVWKVGLLSLHVMILWVVAFCGLALWIFGLLIAQSSPSLHALSV 309

Query: 294 XXXXXXXXXXXMMARTALCHGNCNSIYPPPCKAVPVAFWXXXXXXXXXXFSSGTAYTHSIR 353
 M+AR LHGTG + P + AF L F S +
 25 Sbjct: 310 GSMGLSLILAMIARVLTGHTGRPLQLPAGIIIG-APFL---FNLGTAARVFLSVAWPGGLW 365

Query: 354 TSSVLFALALLVYAWKYIIPWLIRPRSDG384
 ++V + LA +Y W-Y P L+ R D+ PG
 Sbjct: 366 LAAVCWTLAFALYVWRYAPMLVAARVDGHPG 396

30

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 85

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 715>:

35 1 ..ATGCCGCTCTG AAGGTTCAGA CGCGCmTCGGT GyCGGGGAAy CAGAACGyGGT
 51 AGCCGATGCC CAATGAGACT TCGTGGGTT TGAAGCCGGT GTTTTCCAAG
 101 CGTCCCAGT TGTGTAACCG GTATCGCGG TCyAarGTCA GCTTGGGGyGT
 151 GATGTCGAA CCAGCACCCG CGATGACAC AAGACCyAmG CTGTCGATrC
 201 TGTKGCTTGTGATAGGSA GGTTCGyTGC kmksasTyTGC TayrATwkkG
 251 CCTsCwTgKA kAGmGCCtK CkyTGTGKK swGrAr-TAG TCGTGGTTT
 301 T'kTyCACC GAATGACyT GATGTTAACG GTGTCGCTAG GCGACGGGG
 351 CCCCGATATA GGGTTTGAAT TTATCCTGTA GTTTCGAAATC GFAAATGGGG
 401 GACAAAGCCGA GAGAGAAAC GCGCTGGAG CTGCGGTTC CCTGATGTTT
 451 TGTTCGGGT TCTTGTAGT TGTGTTAT CTCTTCAGTA ACTTTTTAG
 501 TAGAGAAATT ACTTTCCTTC CATTTCCTGT AACTGGCAT AACTCGCCGCT
 551 ATTCCTCCAC CGCCGAAATC ..

This corresponds to the amino acid sequence <SEQ ID 716; ORF67>:

1 ..MPSEGSDGXG XGEXEXVAHA QXDFVGFEG VFGQASPVVVT VSGVXXQLGX
 51 DVETDTGDGT KTXAAXDVAF VYGRFEXGXXL YXXAXXXAX XWXXXXSRGF
 50 101 XXHMRNLMFN VSVDGARADI GFFEFIVEFII VNGGQAERRN GVEAAVSIMF
 151 CLGFFVVVVY LFSNFFPSRRI TFFFPESTVGI ICRYSPAAEI ..

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.gonorrhoeae*

ORF67 shows 51.8% identity over 199 aa overlap with a predicted ORF (ORF67ng) from

55 *N.gonorrhoeae*:

-409-

	crf67.pep		MPSEGSQDSGXGXGEKEXXVAHAAQDXFVGFEAG	30
	crf67ng	TNFEIAVLGSMTVRVVFYCARPAPVNGRLLKMPSEGSDGIGIGBSEVAHAAQRQDFGVGFEAG		146
		90 100 110 120 130 140		
5	crf67.pep	VFQASPVVVTVSGVXXQLGXDVETDTGDDTKTXAADKVAFVIGRFXGXXLYXXXXXXXXAX		90
		:: : : : : : : : : : : :		
	crf67ng	VFQASPVVVAVAGVQQAGROVYAHARHRAEAAQAAAVALIFGVLPRMSVRINRNCCWSI		206
10	crf67.pep	XWXXXRSRGFXXXHMRMLMENVSVGDARADIGFFPEVFEIVNGQAERRNGVEAASLME		150
		: : : : : : : : : : : : : : :		
	crf67ng	TRVGGKSTCYFFSIRDAVDSVSGDARYDIGFFFVFEIVNGQAERRNGVECAVFLMF		266
15	crf67.pep	CLGFFVVV-----VYLFNSNFFSRRITFF--PFSVTGIIICRYSPAEEI		190
		: : : : : : : : : : :		
	crf67ng	RLLVYFKVLAALKSFIILSPLDFYVHGIFTIVVFFFTGIGIRGDAAPAEVVADRARHPVGDM		326

The ORF67ng nucleotide sequence <SEQ ID 717> is predicted to encode a protein comprising amino acid sequence <SEQ ID 718>:

20 1 MPSETGSIVV NGVDESGVE SPFFPSIQTQ YRFHIIHRIH RFLPGGMF
51 NRHSIGSGNL GRGVATWVLIS DFKPGCVQF PAGACMTNGE IAVLSGTMVR
101 2 VFYCARPAPV NGRGLKRMPSF GSDGIGIGES EVAHAAHQGF FEGVAFQAG
151 SPVVAVAGV QGQKARGDYVA HARHRAQAQ ARAAFLVFG FLRMSVRINR
201 NCCVSITRVC GKSTCYVFV DADTSVDSG DARIDTGFEE VFEEVIFNG
25 1 QAERRNNGVEC AVFLMLFLVV FTVKLVAKS FILSPLDFV HVGIFVIFPF
301 2 PTVGIIRGDA PAAEVADRH PGVDGMRDVT SEIIAYRAYF VFAWSGWFR
351 IVGNAFGGVB *

Based on the presence of a several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

30 Example 86

The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 719>

1 ATTTGGCTT TTAGAACG CTTTTGGTC ATTACCGGGT AAGCGCTT
 51 TTTTTGTAT TTGGTCATC GCGGTTTGG CGTACCGGT CCGGAGGATT
 101 TGRCCTTGGT AACAGCGGC GTGATTTCGG TGTATGGTTA TACCAATCCG
 151 CATATTATGT TTGAGCTGG TAGTGCTGGC GTATGGTGG CAGGCGCAT
 201 CATGGTCCC GGCGACGAGA TTGGGGCCA CATGATTCCT EGGTCAAC
 251 CTATIGGC CATCACGAGC CGGFAACGGT ATGAGGAGGT TCGAAAAGA
 301 TTGCACAAAGG CGCTTAACG GGTCTTATC FGGCCCGTC TTCTGGCCCG
 351 TTGGAGAACG GCGCTTATG TTACAGCGGG TATCGACGCC AAGGTTTCA
 401 ACTTGGCTT TATCATTAAG GATGGCTGG CGCGA...

This corresponds to the amino acid sequence <SEQ ID 720; ORF78>:

1 MFAFLAFFF EYGYAAVFFV LVIQCFGVP I PEDLTVTG G VISGMGTYNP
51 HIMFAVGMLG VLVGDGIMFA AGRIGWQXXL XFXPIAXIMT PXRYEQVQEK
101 FDKYGNWLF VARFLPGLRT AVFVTAGLSR KVSYLRFIIIM DGLAA...

45 Further work revealed the complete nucleotide sequence <SEQ ID 721>:

1	ATGTTGGCTT	TTTAGAACG	CTTGTTCGTC	AAATGACCGAT	ATTCGCGGAT	ATCCGGGATTC
51	TTTTTTTTTA	TTGGCTTCA	CGCGGTTCGG	CTGGCGGATTT	GGGGAGGATTTC	GGGGAGGATTTC
101	TGACCTTGTG	AAACAGCGGC	GTCGATTTGGC	GTATGGTTCA	TACATCCCGG	TACATCCCGG
151	CATATTAGTG	TTGCGATTCG	TAATGCTCGGC	GTATGGCTGA	GGGGAGGATTTC	GGGGAGGATTTC
201	CATGTTGCCG	GGGCCGAGGG	TTGGGGGCGA	AAAAAACTTA	AGGTAAACCA	TCAGGAAAAGAA
251	CATTTGGCG	CATCATGAGC	CGGAAAGCTT	ATAGCAGGCT	TCAGGAAAAGAA	TCAGGAAAAGAA
301	TTCCSACRAAT	ACGGCTTAAT	GGTCTTATTG	GTTCGGCGCT	TCCTGGCCG	TCCTGGCCG
351	TTTGGACAAACG	GGCGCTTATTG	ATTCAGCGGC	ATTCAGCGGC	AAAGTCATTC	AAAGTCATTC
401	ACTTGGGTTT	TATCATTTAG	GAGGACTGCTG	CGCGACTGTAT	TTCCGGGATTC	TTCCGGGATTC
451	ATTTGGGATT	ATTCGGCGCA	ATACGGCTCC	CAACACATG	ATTCGGCTGA	ATTCGGCTGA

-410-

501 GGCAGRAAATG CACRGCCCTGC AATCGGGTAT TTTTGTATC TTGGGTATAG
 551 GTGCGACCGT TGTGCGCTTG ATTGGTGTGA AAAAACGCCA ACCTATCCAG
 601 TTTCACCGCA GCAATTGGA AGAAAGGGG CGCAACGCA AGCCGCCAA
 651 GGCAGCCAA AAGCCGCGC AAAGCAAACAA ATA

5 This corresponds to the amino acid sequence <SEQ ID 722; ORF78-1>:

1 MFAFLAEFFV EYGYAAVFFF LVICGFGVPI PEDLTLVTGG VISMGMGYTNP
 51 HIMFAVGMLG LVLGDGIMPF AGRIGWGKIL RPKIARIMT PRRYQVQEKK
 101 FDKGNGNWLF VARFLPGLP AVFTAGISR KVSYLRFIIM DGLAALISVP
 151 IWIYLGEGYA HNIQWLMAKM HSLQSGIFVI LGICATVVAW IWWKKRQRQIQ
 10 201 FYRSKLKEKR AQRKAAKAAK KAAQSKQ*

Computer analysis of this amino acid sequence predicts several transmembrane domains, and also gave the following results:

Homology with the dedA homologue of *H.influenzae* (accession number P45280)

ORF78 and the dedA homologue show 58% aa identity in 144aa overlap:

15 Orf78: 4 FLEAFFFEYGYAAVFFFV LVICGFGVPIPEDLTIVTGGVISMGMGYTNPHIMFAVGMLGV 61
 FL FF EYGV AV FVL-ICGFGVPIPED-TLV+GIV+G- N H-M V M-GV
 DedA: 20 FLIGGFTTEYGVWAVLFLVIIICGFGVPIPEDLTIVSGVGIAGLYPENVNSHMLLVSMTG 79

20 Orf78: 62 LVVGDGIMFAAGRIWGQGXKXLXFPIAXIMTPXRYEQQVQEKFQKYGNWLFVARFLPGLRTA 121
 LD GD M+ GRI+G L F PI I+T R V+EKF +YGN VLFVARFLPGLR
 DedA: 80 LAGDSMCYWLGRIGYGTKLRLFRPRIRIVTLQLRMVRFQYGNRVLFVARFLPGLRAP 139

25 Orf78: 122 VEVFTAGISRKVSYLRFIIMDGLAA 145
 +++ +GI+R+VSY+R+I++D AA
 DedA: 140 IYMSVGITRVSYVRFVLIIDFCAA 163

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF78 shows 93.8% identity over a 145aa overlap with an ORF (ORF78a) from strain A of *N.meningitidis*:

	10	20	30	40	50	60
30 orf78.pep	MFAFLAEFFFEYGYAAVFFFV	LVICGFGVPIPEDLTIVTGGVISMGMGYTNPHIMFAVGMLGV				
orf78a	MFAFLAEFFFEYGYAAVFFFV	LVICGFGVPIPEDLTIVTGGVISMGMGYTNPHIMFAVGMLGV				
	10	20	30	40	50	60
35 orf78.pep	VLVGDGIMFAAGRIWGQGXKXLXFPIAXIMTPXRYEQQVQEKFQKYGNWLFVARFLPGLRTA					
orf78a	VLVGDGIMFAAGRIWGQKLKRPRIARIMTPKRYAQQVQEKFQKYGNWLFVARFLPGLRTA					
	70	80	90	100	110	120
40 orf78.pep	AVFVFTAGISRKVSYLRFIIMDGLAA					
orf78a	AVFVFTAGISRKVSYLRFIIMDGLAA					
	130	140				
45 orf78.pep						
orf78a						
	130	140	150	160	170	180

The complete length ORF78a nucleotide sequence <SEQ ID 723> is:

1 ATGTTGGCCC TTTTGGAAAGC CTTTTTGTC GAATACCGGT ATGCGGCCGT
 51 GTTGGTTCTGT TTGGTCATCT GCCTGGTTGGG CCTCGCCGATT CCGGAGGATT
 101 TGACCTTGCT AACAGCGGC GTGATTTCGG GTATGGTTTA TACCAACTCG
 151 CATATTTATGT TTGGCATGCG TATGCTCGGC GTATGGTGG GGGACGGCAT
 201 CTGGTCCCC CGCCGGAGCA TCTGGGGCCA GAAAATCTC AGTCTCACAC
 251 CGATTGCGCG CATCATGAGC CGGAAACGTT ACGGCACAGGT TCAGGAAAAAA
 301 TTGGCAAAATG AGCGGCAACTG GGTTGTTATGT GTGCGCTCGT TTCTGCCCCG
 351 TTGGCGGACT GCCGGTTCTGG TTACCGCCGG CATCAGCGGC AAAGTATCGT
 401 ATCTGGCTT CTCGATTATG GACGGCGCTTG CGGGCGCTGT TTCCGTCGCC
 451 GTTGGATT ACTGGGCGA GTACGGCGCG CACAAACATCG ATGGCGTGT

501 GGCAGAAAATG CACRGCCCTGC AATCCGGCAT CTTCATCGCA TTGGCGGTGC
 551 TGGCGCCGCG GCTGGCGTGG TCTGGGTGC GCAAACCGG ACATTATCAG
 601 CTTTACCGCG CACATTGCG CGAAAAACGC GCCAAACGCA AGGGCGAAAG
 651 GGCAGCGAA AAAGCGCAC AGNAGCAGCA GTAA

5 This encodes a protein having amino acid sequence <SEQ ID 724>:

1 MFALLEAFFV EYGYAAVFFFV LVIICGFGVP I PEDLTLVTGG VISMGMYTNP
 5 HIMFAVGMLD LVLGDGIMPF AGRINGKQIL KFKPIARIMT PKRYAQVQEKG
 10 FDKYGNWLF VARFLPLRT AVFVTAGISR KVSYLRFLIM DGLAALISVP
 15 WIYLGEYGA HNIDWLMARM HSLQSGIFIA LGVLAAAALW FWWRKRRHYQ
 201 LYRAQLSEKR AKRKEAKAAK KAAQKQQ*

ORF78a and ORF78-1 show 89.0% identity in 227 aa overlap:

	10	20	30	40	50	60
15	orf78a.pep	MFALLEAFFV EYGYAAVFFFV LVIICGFGVP I PEDLTLVTGG VISMGMYTNP :: : : : : orf78-1	MFALLEAFFV EYGYAAVFFFV LVIICGFGVP I PEDLTLVTGG VISMGMYTNP HIMFAVGMLD :: : : : : 10 20 30 40 50 60			
	70	80	90	100	110	120
20	orf78a.pep	VLVGDGIMFAAGRIWGKILKFKPIARIMTPKRYAQVQEKFDKYGNWLFVARFLPLRT :: : : : : orf78-1	VLVGDGIMFAAGRIWGKILKFKPIARIMTPKRYAQVQEKFDKYGNWLFVARFLPLRT :: : : : : 70 80 90 100 110 120			
	130	140	150	160	170	180
25	orf78a.pep	AVFVTAGISRKVSYLRFLIMDGLAALIISPVWVWIYLGEYGAHNIDWLMAKMHSLSQSGIFIA :: : : : : orf78-1	AVFVTAGISRKVSYLRFLIMDGLAALIISPVWVWIYLGEYGAHNIDWLMAKMHSLSQSGIFIA :: : : : : 130 140 150 160 170 180			
	190	200	210	220		
30	orf78a.pep	LGVLAALAWFWWRKRHHQLYRAQLSEKRKAEEAKKAQKQX :: : : : : orf78-1	LGIATVVAIWKKRORIQFYRSKLKEKRAQRKAEEAKKAQSKQX 190 200 210 220			

Homology with a predicted ORF from *N.gonorrhoeae*

ORF78 shows 97.4% identity over 38 aa overlap with a predicted ORF (ORF78ng) from *N. gonorrhoeae*:

40	orf78.pep	XXLXFXPAXIMTPXRYEQVQEKFDKYGNWVLEVARFLPGLRTAVFVTAGISRKVSYLRF :: : : : : orf78ng	YPVLFVARFLPGLRTAVFVTAGISRKVSYLRF	137 32
45	orf78.pep	IIMDGLAA :: : : orf78ng	IIMDGLAALISPVWVWIYLGEYGAHNIDWLMAKMHSLSQSGIFIALGVLAALAWFWWRKR	145 92

The ORF78ng nucleotide sequence <SEQ ID 725> is predicted to encode a protein comprising amino acid sequence <SEQ ID 726>:

50 1 ..YPVLFVARFL PGLRTAVFVT AGISRKVSYL RFLIMDGLAA LISPVWVWIYL
 55 51 GEYGAHNIDW IMAKMHSLSQG GIFIALGVLA AALAWFWWRK RRHYQLYRAQ
 101 101 LSEKRAKRKA EKAARKKAQK QQ*

Further work revealed the complete gonococcal nucleotide sequence <SEQ ID 727>:

55 1 atgtttggcc tttTggaaagc CTTTTTGTC GAAtacggCt atgcGCCGT
 51 GTTCTTCGTT TTGGTCATCT GCGCTTGTGG CGTCGCCGATT CCCGAGATTT
 101 TGACTCTGGT AACGGGGCGC GTGATTTGCG GTATGGTTTA TACCAATCCG
 151 CATATTAATGTT TTGGCGTGG TAGTCTCGGC GTGTTGGCGG GCGACCGGGT
 201 GATGTTGGCC GCGGGAGCGA CTCTGGGGCA GAAAATCCTC AAGTTCAAAAC
 251 CGATGTCGGG CATCATGAGC CGGAAACGTT AGGGCAGGT TCAGGAAAAAA
 301 TTGCGAACAAAT AGCGGCAACTG GGTTCTGTTT GTGCCGCCGTT TCTCTGCCGG

-412-

351 TTGGGGAAGCT TCGATTATGC TTACGGCCGG CACCGCCGC AAAGTATCTG
 401 ATCTGCCGT TCTGATTATG AGCGGCCGTC CGCCGCTGAT TTCCGGCTGCC
 451 GTTTGGATT ACTTGGGCGA STACGGCGCC CAAACATCG ATTGGCTGAT
 501 GGCAAAATG CACAGCGTC AACTGGCGAT CCTACGCGA TTGGCTGCC
 551 TGGCGCGGC GCTGGCGTGG TTCTGGTGGC GAAACCGG ACATATTAG
 601 CTTACCCCG ACAAATGGC CGAAAACCGC GCAAAACCGA AGGGGGAAAA
 651 GGCAAGCGAA AAAGCGCGAC AGAACAGCGA GTta

This corresponds to the amino acid sequence <SEQ ID 728; ORF78ng-1>:

1 MFALLAEFFV EYGYAFFFVY LVICGFGVTP
 51 HIMFVGMLV VLAGDGVMPA HNGQKQIL KFKPIARMT
 101 FDKGYNWVL VARFLGLRAT AVFVTAGISR KVSYLRFILM DGLAALISVP
 151 WTYLGEYGA HNDWLMRM LHSQGIFTA LGVLAAALAW FWWRKRRHYQ
 201 YLAQSEKR AKRRAEKA KAAQKQO⁹

ORF78ng-1 and ORF78-1 show 88.1% identity in 227 aa overlap:

15		10	20	30	40	50	60
	orf78-1.pep	MFAFLAEEFVYEGYAAFFVFLVICFGVPIPEDLTIVTGGVISGMGTYNPHIMFAVGMLG 					
	orf78ng-1	MFALEAFVEFYEGYAAFFVFLVICFGVPIPEDLTIVTGGVISGMGTYNPHIMFAVGMLG 10 20 30 40 50 60					
20		70	80	90	100	110	120
	orf78-1.pep	VLVGDGVMFAAIGRIWGQKILRFKKP1ARIMTPKRYSQEKFDKYGNWNLVVARFLPGLRT 					
	orf78ng-1	VLAGDGVMFAAIGRIWGQKILRFKKP1ARIMTPKRYSQEKFDKYGNWNLVVARFLPGLRT 70 80 90 100 110 120					
25		130	140	150	160	170	180
	orf78-1.pep	AVFVTAGISRKVSYLRLF1IMDGLAALISVPIWIYLGEYGAHNIDWLMAKMHSLSQSGIFV 					
	orf78ng-1	AVFVTAGISRKVSYLRLF1IMDGLAALISVPIWIYLGEYGAHNIDWLMAKMHSLSQSGIFV 130 140 150 160 170 180					
30		190	200	210	220		
	orf78-1.pep	LGIATVVAVIWWKKRQIOPYRSKLKEKKAQRKAAAKKAQSKQX : : : : : : : : :					
	orf78ng-1	LGVLAAALAWIWWKRKRHYQLYRAQSLKEEAKRKAEEAKKAQKQQX 190 200 210 220					
35							

Furthermore, orf78ng-1 shows homology to the dedA protein from *H. influenzae*.

```

40    sp|P45280|YG29_HAEIN HYPOTHETICAL PROTEIN HI1629 >gi|1073983|pir|D64133 dedA
protein (dedA) homolog - Haemophilus influenzae (strain Rd KW20)
>gi|1574476 (U38236) dedA protein (dedA) [Haemophilus influenzae] Length = 212
Score = 223 bits (563), Expect = 7e-58
Identities = 108/182 (59%), Positives = 140/182 (76%), Gaps = 2/182 (1%)

45    Query: 5 LEAFFVEYGYAAWFVFLVICGFGVPIPEDLTIVTGGVVISGM--GYTNPHIMFAVGMLGVL 62
      L FF EYGY AV FVLICGFGVPIPEDLTIV+GGV+*+ N H+M V M+GVL
      Sbjct: 21 LIGFFTEGYWAVALFVFLIICGFGVPIPEDLTIVTGGVVISGM--GYTNPHIMFAVGMLGVL 80

50    Query: 63 AGDDGVMFAAGRINWQOK1LKFKP1IAR1MPKRYAQLQEKFKDYGNNWLFVARFLPGLRTAV 122
      AGD M+ GRI+G KIL+F+PI RI+T +R V+EKF +YGN VLFLVARFLPGLR +
      Sbjct: 81 AGDSCMYWLGR1YGTKLRLFRPRLIRRIVTLQRLRMVREKFQSQYGNRNLFLVARFLPGLR API 140

55    Query: 123 FVTAGISRKVSYLSRFLLIMGLAISLVPWVIYLGEYGAHNIDWLMKAKMHSLQSGLIFIALG 182
      +++GI+R+VSY+RF++# AA+ISVP+WIYLGE GA N+DWL ++ Q I+I +G
      Sbjct: 141 YMVSGITRVRVSYRVFLIDFCAAIIISVPINYLGEIAGAKNLDLWHLHTQIKQKGQIVIYFIG 200

60    Query: 183 VL 184
      L
      Sbjct: 201 YL 202

```

Based on this analysis, including the presence of putative transmembrane domains, it is predicted that these proteins from *N. meningitidis* and *N. gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 87

- 5 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 729>:

```

1 ATGAAATTAT TATTGGCGGC CGTGTAGTATG CGAACCTTTGG CAGGGGGGGT
51 TCTTCGGCCGC TGGGTCACCG CTGAGGGACGG CTGGGGCGC ACACCGCTC
101 AAGGTATGAA AATAGGCGGC GCCTTCTATGA AAATTCACAA CGAACGAGG
151 AAACAAAGACT TTGGTCCTGG CGGAAAGCAGC CCGCTTGCG ACCGGAGTC
201 AGTGTATCATG CACATCAACG ACACAGCGCT GATTCGGATG CGCGGAGTCG
251 AAGGGCCGGGT GCCTTGGAA CGCAAAATGGT TTACCGAACAT CCGGGAGCCG
301 AGCTATCATG TGATGTTTAT GGTTTTAAAGA AACAACTTAA AAGGGCCGG
351 TAAAATTCG GTTACCTCCG AATTTAAAGG CGCCAAAGCGG CAACCGCTC
401 RACTGGRAAGT AACATTCGGC CGCATGGCG CGATTCGAGA C... .

```

- 15 This corresponds to the amino acid sequence <SEQ ID 730; ORF79>:

1 MKKLLAAMM AGLAGAVSAA GHVEDGWAR TTVEGMKIGG AFMKIHNDDEA
51 KQDFLLGGSS PVADRVEVHT HINDNGVMRM REVEGGVPLE AKSVTELKPG
101 SYHVMFMGLK KQLKEGDKIP VTLKFKNAKA QTVQLEVVKIA PMFAMNH..

Further work revealed the complete nucleotide sequence <SEQ ID 731>:

20	1	ATGAAATTTA	TATTGGCCGC	CTGTGGATGATG	CGAGGTTTGG	CAGGGGGTGT
	51	TTTCGGCGGC	TTACCGCACG	TTGAGGAGCG	TGGGGCGGCC	ACCAAGCTTGT
	101	AAAGGTATGAA	ATAAGCCGCG	GCTGTGATGA	AACTCCACAA	CGAACGAGGCG
	151	AAACAAAGCT	TTTTGCTCGG	CGGAAGCAGC	CCCTGGTCG	ACCGGGTGCGG
	201	AGTCGTACAC	CAACATACAG	AACAGCGGT	GATGGCGGATG	CGCGAGTCGT
25	251	AAAGCCGGGT	GCTCTTGGAA	GGAAATTCG	TTTACCGAATG	CAACACGGGCG
	301	AGCTGATCATG	TGATGTTATG	GGGGTTTAA	AAACARTTAA	AAAGGGCGGCG
	351	TTAAATTCATG	GTATTCCTGA	ATTTTAAAGA	CGGAACTTAA	CAACAGCTCG
	401	AACTGGAAAGT	AAACATTCATG	CGCATGCGG	CAATGAACCA	CGGTCTACG
	451	CAOGCGGAG&	GGCTACGCGA	CTAA		

- This corresponds to the amino acid sequence <SEQ ID 732: ORF79-1>:

1 MKKKLAAVMM AGLAGAVSA GVHVEDGWAR TTVEGMKIGG AFMKIHNDEA
51 KQDFLLGGSS FVADRVEVHT HINDNGVMRM REVEGGVPLE AKSVTELKPGH
101 SYHVMFMGLK KQLKEGDKIP VTLKFKNAKA QTVQLEVKIA PMPAMNHGH
151 HCPGRH*

- 35 Computer analysis of this amino acid sequence revealed a putative leader peptide and also gave the following results:

Homology with a predicted ORF from *N. meningitidis* (strain A)

ORF79 shows 94.6% identity over a 147aa overlap with an ORF (ORF79a) from strain A of *N. meningitidis*.

		10	20	30	40	50	60	
40	orf79 pep	MKKLLAAVMAGLAGRVSAAVGHVVEDGWARTVTEGMKIGGFMKIHNEAKQDFLGGSS						
	orf79a	MKKLLAAVMAGLAGRVSAAGIGHVEDGWARTVTEGMKIGGFMKIHNEAKQDFLGGSS						
45		10	20	30	40	50	60	
	orf79 pep	PVADRVVEVHHTINDNGVMMRMEVEGGVPLEAKSPTTELKGPGSYHVNFMGLKQLKEKGKIE	70	80	90	100	110	120
50	orf79a	PVADRVVEVHHTINDNGVMMRMEVEGGVPLEAKSPTTELKGPGSYHVNFMGLKQLKEKGKIE	70	80	90	100	110	120

		130	140
	orf79.pep	VTLKFKNAKAQTVQLEVKAIPMPAMNH	
5	orf79a	VTLKFKNAKAQTVQLEVKAIPMSAMDGHHHGEAHQHX	
		130	140
			150

The complete length ORF79a nucleotide sequence <SEQ ID 733> is:

1	ATGAAANNAAAC	TATGGCCAG	CGTGTGATGATG	CGAGGGCTGG	GGGGGGGTGG
5	TTCGGGCC	GGATTCAGCG	TTGAAGCAGCG	CGAGGGCTGG	ACCCACCTCG
10	101 AAGGTATGAA	ATAAGGGCCC	GCGCTTATGATG	AAATCCCAAA	CGAAAGGGAC
15	151 AAACAAGACT	TTTGTGTCG	CGGAAGGACCG	CTGTGTCGCG	CCCGCTGCGA
20	201 ATGTCATGCA	CATATCATAG	ATTAACGGTGT	GTATGCAGTG	CGGCAAGTCG
25	251 AAGGGCGGCG	GCCCTTGGAA	GCGAAATTCG	TTCAGAACAT	CAACCGGGC
30	301 AGCTATGATC	TCATGTTGAT	GCGTNTGAAA	AAACAAATTGA	AGANGGCGCA
35	351 CAAGATTC	GTACCTCTGA	ATTTTGTAAA	CGCCAAAGTC	CAACCGTCG
40	401 AACTGGAAGT	AAACCCGGAG	CGGATGTGTCG	CAATGGACCA	CGGTCACTAC
45	451 CAGCGGAGA	GGCTCATCAGCA	CTAA		

This encodes a protein having amino acid sequence <SEQ ID 734>:

20 1 MKXLLAAVMM AGLAGAVSA GIHVEDGWAR TTVEGMKMG AFMFIHNDEA
51 KQDFLGGSS PVADRVEVHT HINDNGVMRN REVEGGVPLE AKSVTELKPG
101 SYHVMFMGXK KQLKXGDKIP VTLKFKNAKA QTVOQLVEKTA PMSAMDGHHH
151 HGEEAHQH*

ORF79a and ORF79-1 show 94.9% identity in 157 aa overlap:

Homology with a predicted ORF from *N. gonorrhoeae*

ORF79 shows 96.1% identity over 76 aa overlap with a predicted ORF (ORF79ng) from *N.gonorrhoeae*:

45	orf79.pep	FMKIHNDEAKQDFLLGGSSPVADRVEVHTHINONGVMRMREVEKGVPLEAKSVTELKPGS	101
	orf79ng	INDNGVMRMREVEKGVPLEAKSVTELKPGS	30
50	orf79..pep	YHVMFMGLKKQLKEGDKIPVTLKFKNAQTVQLEVKIAPPMAMNH	147

An ORF79ng nucleotide sequence <SEQ ID 735> was predicted to encode a protein comprising amino acid sequence <SEQ ID 736>.

1 ..INDNGVMRMR EVKGGVPLEA KSVTELKPGS YHVMFMGLKK QLKEGDKIPV
51 TLKFKNAKAQ TVQLEVKTAP MSAMNHGHHH GEAHQH*

Further work revealed the complete gonococcal DNA sequence <SEQ ID 737>;

1	ATGAAAAAAAT	TATTGGCAGC	CCTGATGATG	GCAGGTTTGG	CAGGCAGGTT
5	51	TTccggcCc	GGatTccATG	TGAAggACGGC	CTGGGGCGC accaTGtcg
101	aaggATGaa	aattggCGGC	CGatTCATga	aaATCCACAA	CGACaaGcc
151	atacaAGACT	tttgTgcTcg	CGGaaagatg	cggatgtgcgg	acgcgSTCGA
201	AGTGCAtaca	caCATCAAC	ACAAGCGGT	GATGCGTATG	CGCGAATGTC
251	AAGGGCGGT	GCTTTGGAG	GGAAATCCG	TTACCGAACT	CAAAACCGGC
301	AGCTTACAG	TGATGTTTA	GGGTTGAAR	AAACRACRTGA	AGAGAGGGCGA
351	CRAGATTCCC	GTACCCCTGA	AAATTTAAA	GGCCAAAGCG	CAAACCGTCC
401	ACTGGAAGT	CAAAACCGCG	CGGATGTCGG	CAATGAAACCA	CGGTCACTAC
451	CACGGCGAAG	CGCATCAGCA	CTTAA		

This corresponds to the amino acid sequence <SEQ ID 738; ORF79ng-1>:

1	MNKLLAAVMAGLAGAVSAA	GVHVEDGWAR	TTVEGMKMGG	AEFKIHNDNE
5	101	IQDFVILGGSM	PVADRVEVHT	HINDNGVMRM
15	151	SYHVMEMGLK	QLKEGDKIP	QTVQLEVTKTA
		HGEAHQR*		PMSAMNHGHHH

ORF79ng-1 and ORF79-1 show 95.5% identity in 157 aa overlap:

10	10	20	30	40	50	60	
20	orf79-1.pep	MNKLLAAVMAGLAGAVSAA	GVHVEDGWAR	TTVEGMKMGG	AEFKIHNDNE		
10	orf79ng-1	MNKLLAAVMAGLAGAVSAA	GVHVEDGWAR	TTVEGMKMGG	AEFKIHNDNE		
20		10	20	30	40	50	60
25	orf79-1.pep	PVADRVEVHT	HINDNGVMRMRE	VEGGVPLEAKSVTELKPGSYHV	MFGLKKQLKEGDKIP		
10	orf79ng-1	PVADRVEVHT	HINDNGVMRMRE	VGKGGVPLEAKSVTELKPGSYHV	MFGLKKQLKEGDKIP		
		70	80	90	100	110	120
30	orf79-1.pep	VTLKFNAKAQTQV	LEVKIA	PAMPAMNHGH	HHGEAHQHX		
10	orf79ng-1	VTLKFNAKAQTQV	QLEVKTAPMSAMNHGH	HHGEAHQHX			
		130	140	150			

Furthermore, ORF79ng-1 shows significant homology to a protein from *Aquifex aeolicus*:

35	gi 2983695	(AE000731)	putative protein	[<i>Aquifex aeolicus</i>]	Length = 151
	Score = 63.6 bits	[152],	Expect = 6e-10		
	Identities = 38/114	(33%),	Positives = 58/114	(50%),	Gaps = 1/114 (0%)
40	Query: 24	VEDGWARTTVE	VEGMKMGAFN	KIHNDNEAIQDVF	LGGSMPPVADRVEVHT
	V+ W	G	M I N+	D++G	+A RVE+H + +N V +M
	Sbjct: 27	VKHFWVMEPPPG	PNTTMMGMIIV	NGDEFDYLIGAKTDI	QKVLELHKTVIENDVAKMVPQ
45	Query: 84	KGGVPLEAKSVTEL	KPGSYHV	MFGLKKQLKEGD	KIPVTLKFNAKAQTQV
	+++ K	E K	YHV	MGK	LEV 137
	Sbjct: 67	ER-IEIPIPKGKV	EHHGHYHV	MIIGLRRKRI	KEGDKVKEFKGSKITVEAPV
					139

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF79-1 (15.6kDa) was cloned in the pET vector and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 18A shows the results of affinity purification of the His-fusion protein. Purified His-fusion protein was used to immunise mice, whose sera were used for ELISA (positive result) and FACS analysis (Figure 18B). These experiments confirm that ORF79-1 is a surface-exposed protein, and that it is a useful immunogen.

Example 88

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 739>:

```

5      1 ATGACGGTAA CTGGCGCCGA AGGC GGCAAA GCTGCCAAGG CGTTAAAAAA
51     51 ATATCTGATT ACGGGCATTT TGGTCTGCCT GCCTGATTCGG GTTACCGGTT
101    101 GGGTTGTTTC CTATATCGTT TCCGCGTCCG ATCAGCTCGT CAACTCGCTG
151    151 CGGAAGCAAT GGCGCGCCGA ATATGTTTTG GGGTTTAATA TCCC CGGGCT
201    201 GGGCGTTATC TTGTCGATTC CGGTATTTGT TGTAACCGGA TTGTTTGCCTG
251    251 CCAACGTTTGGTCCGCAATCCTCGCCCGTGGGACAGC CTGTTGTTGGGG
301    301 CGGATTCGG TTGTTGAAAC CATCTATTCC AGTGTGAAAAA AGATATCCGA
351    351 ATAcTGTGCTTCCCGACAGCGCCGATCTGCTT TAAACCGGGC GTTACTCGTGC
401    401 CGTTCCCTCA GCGCGGTTATC TTGAGCAGTyy CTTTCGTC AGGGCAGGTG
451    451 TCGAATTCGG TTAAAGCGCA ATTCGGCAAs GACGGCGGATT ATCTTTCCGT
501    501 GTATGTTCCG ACCACGCCGA ATCCGACCGG CGGTATTTAT ATTATGGTAA
551    551 AGAAAAGCGA TTGTCGCGGA CTGCGATAAGC CGGTGGGACGA AsCATTGGAA
601    601 TATGTTGATT CGCTGGGTATC CTGCGATAAGC CGGTGGGACGA AsCATTGGAA
651    651 ATTGGCAGSGA CCTATGCGT CTGAAAAGGC GGATTTGCCG GAAACAAACAT
701    701 AA

```

This corresponds to the amino acid sequence <SEQ ID 740; ORF98>:

```

20     1 MTVTAAEGGK AAKALKKYLI TGILWLPIA VTVVVVSYIV SASDQLVNLL
51     51 PKQWRPQYVL GFNIPGLGVI VAIAVLFTVG LFAANVLGRQ ILAAWDSL LG
101    101 RIFVVKSIYS SVKRVSIYS SDSSRSFKTP VLVPFFQPGI WTIAFVSGQV
151    151 SNAVKALPLX DDGLYLSVYYP TTNPNTGGYY INVVKSDVRE LDMSVDEXLK
201    201 YVSLGMVIP DDLFVKTIAK PMFSEKADLP EQQ*

```

25 Further work revealed the complete nucleotide sequence <SEQ ID 741>:

```

1      1 ATGACGGTAA nTGCGGCCGA AGGC GGCAAA GCTGCCAAGG CGTTAAAAAA
5      51 ATATCTGATT ACGGGCATTT TGGTCTGCCT GCCTGATTCGG GTTACCGGTT
101    101 GGGTTGTTTC CTATATCGTT TCCGCGTCCG ATCAGCTCGT CAACTCGCTG
151    151 CGGAAGCAAT GGCGCGCCGA ATATGTTTTG GGGTTTAATA TCCC CGGGCT
201    201 GGGCGTTATC TTGTCGATTC CGGTATTTGT TGTAACCGGA TTGTTTGCCTG
251    251 CCAACGTTTGGTCCGCAATCCTCGCCCGTGGGACAGC CTGTTGTTGGGG
301    301 CGGATTCGG TTGTTGAAAC CATCTATTCC AGTGTGAAAAA AGATATCCGA
351    351 ATOGCTGCTTCCCGACAGGA CGCTGGTCTT TAAACCGGGC GTTACTCGTGC
401    401 CGTTCCCTCA GCGCGGTTATC TTGAGCAGTyy CTTTCGTC AGGGCAGGTG
451    451 TCGAATTCGG TTAAAGCGCA ATTCGGCAAs GACGGCGGATT ATCTTTCCGT
501    501 GTATGTTCCG ACCACGCCGA ATCCGACCGG CGGTATTTAT ATTATGGTAA
551    551 AGAAAAGCGA TTGTCGCGGA CTGCGATAAGC CGGTGGGACGA AsCATTGGAA
601    601 TATGTTGATT CGCTGGGTATC CTGCGATAAGC CGGTGGGACGA AsCATTGGAA
651    651 ATTGGCAGSGA CCTATGCGT CTGAAAAGGC GGATTTGCCG GAAACAAACAT
701    701 AA

```

This corresponds to the amino acid sequence <SEQ ID 742; ORF98-1>:

```

1      1 MTEXAAEGGK AAKALKKYLI TGILWLPIA VTVVVVSYIV SASDQLVNLL
51     51 PKQWRPQYVL GFNIPGLGVI VAIAVLFTVG LFAANVLGRQ ILAAWDSL LG
101    101 RIFVVKSIYS SVKRVSIYS SDSSRSFKTP VLVPFFQPGI WTIAFVSGQV
151    151 SNAVKALPLX DDGLYLSVYYP TTNPNTGGYY INVVKSDVRE LDMSVDEXLK
201    201 YVSLGMVIP DDLFVKTIAK PMFSEKADLP EQQ*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF98 shows 96.1% identity over a 233aa overlap with an ORF (ORF98a) from strain A of *N.*

50 *meningitidis*:

	10	20	30	40	50	60
orf98.pep	MTVTAAEGGKAAKALKKYLTITGILWLPIAVTVVVVSYIVSASDQLVNLLPKQWRPQYVL					
orf98a	MTEPAAEGGKAAKALKKYLTITGILWLPIAVTVVVVSYIVSASDQLVNLLPKQWRPQYVL	10	20	30	40	50
						60

55

		70	80	90	100	110	120
5	orf98.pep	GNIPICLGIVIAVLFGLPAAANVLGRQIILAAWDSLGLRIPVVKSIYSSVKKVSEVNL					
	orf98a	GNIPICLGIVIAVLFGLPAAANVLGRQIILAAWDSLGLRIPVVKSIYSSVKKVSEVNL	70	80	90	100	110
10	orf98.pep						
	orf98a	SDSSRSFKTFLVLPFPQGWTIAFVGSGVSNAVKAALPDXGDDYLSVYVPTPNPTGYY	130	140	150	160	170
15	orf98.pep	SDSSRSFKTFLVLPFPQGWTIAFVGSGVSNAVKAALPDXGDDYLSVYVPTPNPTGYY	130	140	150	160	170
	orf98a	IMVKKSVDRELDMGSVDEXLKYVISLGMVIPDDLPVKTLAXPMPSEKADLPSEQQX	190	200	210	220	230
		IMVKKSVDRELDMGSVDEXLKYVISLGMVIPDDLPVKTLAXPMPSEKADLPSEQQX	190	200	210	220	230

The complete length ORF98a nucleotide sequence <SEQ ID 743> is:

20	1	ATGACGCCAAC	CTGGGGCCCA	AGGCCGAA	GCTGCCAAGG	GTAAAAAA
	51	ATATCTGATT	ACGGCGATT	TGGCTTGCCT	GCCGATTCGG	TAACGGTT
	101	GGGGTTGTTCT	CTATATCGT	TCGGCGCCG	ATCAGCTTG	CAACTCTGCT
	151	CGGAAGACCAT	GGCGCGCCGA	ATATGTTTTG	GTTTTATAA	TCGCCGGGT
	201	GGGGCTTATC	GTGGCCATTG	CCTGGTATG	TCTAACCGGA	TTATTTGCG
25	251	CAACAACTATT	GGGGCGCCGA	ATTTCTGCCG	CTGGGAGACG	CTTGTGCG
	301	CGGATTCGGG	TTGGTGAAGTC	CTATCTTATC	ACTGTGAAAAG	ATCTATTCGA
	351	NTCTGGTGTG	CTTACACGAA	GCGCTTCTGT	TTAACACCGA	GTACTCTGC
	401	CTTCCCCCCC	ATTCGGTATT	TGGACATTCG	CTATCGTC	CGTCGATG
	451	TCTGATGGCG	TTAAAGCCCG	ATTCGGCGAG	GAGCGGGAATT	ATTTCTCC
	501	GTATGTCCTC	ACCAACCGCA	GGCTACCGGG	CTGTTACATAT	ATATGGTAA
	551	AGAAAAGCGG	TGTCCGGCA	CTGATCTGAT	GCGTGGAGCA	AGGCTTGA
	601	TATGTGATT	CGCTGGGTAT	GTCATCTCC	GAGGACTCGC	CGCTAACAA
	651	ATTGGCAGGA	CCTATGCCCT	CTGAAAGGC	GGATTTGCC	GAACACAAT
	701	AA				

35 This encodes a protein having amino acid sequence <SEQ ID 744>:

```

1 MTEPAEAGK AAKALKYLYI TGILWLPFIA VTVVVVSYIV SASDQLVNLL
51 PKQWRHQVYF GFLNPGLGVI VAIAVLTFGT LEEAFLWLRQI LIAALMSLLG
101 RPIVKVSIYSV SVKKVKSXLSS DSDRSFKTPE VLFVPPQPSI WTIAFVSGC
151 SNAVKAALEPL DDGYLSVVPV TTNPTGNGYY IMVKHSVDR LEQMSVDEALK
201 YU1SICMUIJR DDLBWTKLDP PMPSKAKLDP EOO*

```

ORF98a and ORF98-1 show 98.7% identity in 233 aa overlap:

		10	20	30	40	50	60
45	orf98a.pep	MTEPAEAGGKAALKYLLTGTILWLPPIAVTVWWVSYIVSASDQLVNLLPKQWRPQQYVL 					
	orf98-1	MTEXAAEGGKAALKYLLTGTILWLPPIAVTVWWVSYIVSASDQLVNLLPKQWRPQQYVL 10 20 30 40 50 60					
50	orf98a.pep	70 80 90 100 110 120 GENIPGLGVIVAIAVLVTGLFAANVLGRQILAAWDSLLGRIPVVKSIYSVKKVSSSL					
	orf98-1	70 80 90 100 110 120 GENIPGLGVIVAIAVLVTGLFAANVLGRQILAAWDSLLGRIPVVKSIYSVKKVSSSL					
55	orf98a.pep	130 140 150 160 170 180 SDSSRSFKTPFLVPFPQSGIWTIAFVGSCQVSNAVKAALPKDGDYLSQLSYVPPTPNPTCGYY 					
	orf98-1	130 140 150 160 170 180 SDSSRSFKTPFLVPFPQSGIWTIAFVGSCQVSNAVKAALPKDGDYLSQLSYVPPTPNPTCGYY 130 140 150 160 170 180					
60	orf98a.pep	190 200 210 220 230 IMVKKSDVRELDMSVDEALKYVISLGWVIPEDDLPVKTLAGCPMPSEKADLPEEQX 					
	orf98-1	190 200 210 220 230 IMVKKSDVRELDMSVDEALKYVISLGWVIPEDDLPVKTLAGCPMPSEKADLPEEQX 190 200 210 220 230					

Homology with a predicted ORF from *N.gonorrhoeae*

ORF98 shows 95.3% identity over a 233 aa overlap with a predicted ORF (ORF98ng) from *N.gonorrhoeae*:

		10	20	30	40	50	60	
5	orf98.pep	MTVTAEEGGKAAKALKYLITGILWLPPIAVTVWWVSYIVSASDQLVNLLPKQWRQYVL						60
	orf98ng	MTPEAAEKGKAAKALKYLITGILWLPPIAVTVWWVSYIVSASDQLVNLLPKQWRQYVL						60
10	orf98.pep	GFNIPGLGVIVIAVLFVTGLFAANVLGRQIILAANDSLLGRIPVVKSIYSSVKVSEYVL						120
	orf98ng	GFNIPGLGVIVIAVLFVTGLFAANVLGRQIILAANDSLLGRIPVVKSIYSSVKVSEYVL						120
15	orf98.pep	SDSSRSFKTPVLPVPQPQGIWTTIAFVGQGSNAVKAAALPXDGDYLSYYVPTTPNPTGGY						180
	orf98ng	SDSSRSFKTPVLPVPQPQGIWTTIAFVGQGSNAVKAAALPXDGDYLSYYVPTTPNPTGGY						180
20	orf98.pep	IMVKKSVDRELDSVDEALKYVISLGMVTFDDLPVKTLAXPMPSKADLPEQQ						233
	orf98ng	IMVKKSVDRELDSVDEALKYVISLGMVTFDDLPVKTLAGPMPPEKAELPEQQ						233

20 The complete length ORF98ng nucleotide sequence <SEQ ID 745> is predicted to encode a protein having amino acid sequence <SEQ ID 746>:

1	MTEPAAEKGK AAKALKYKL TGLWLPPIA VTVWWVSYIV SASDQLVNLL
51	PKQWRQYVL GNFNIPGLGV VAIAVLFVTG LFAANVLGRQ IILAANDSLLX
101	RIPVVKSIYS SVVKVSESSL SDSSRSFKTP VLVPFPQSGI WTIAFVSGOV
151	SNAVKAAALPQ DGDYLSYYVP TTNPNTGGY IMVKKSVDRE LDMSVDEALK
201	YVISLGMVIP DDLPVKTLAG PMPPEKAELPEQQ*

Further work revealed the complete nucleotide sequence <SEQ ID 747>:

1	ATGACGGAAC CTGGCGGCCA AGGCCGCAA GCTGCCAAGG CGTTAAAAAA
51	ATATCTGTAC ACAGCGATT TGGCTGGCTT GCGCATTTGGC GTAACCGGTT
101	GGGGTTTCCTT CTATATCGTT TCCCGCTCC ACCAGCTTGT CAACCTGCTG
151	CCGGACGACAT GGCGCCGCCA ATATGTTTTG GGGTTTAATA TCCCCGGGT
201	CCGGCTGTATT GTTGCCATTG CCGTAGTTGT TGTAAACCGGA TTATTCGCG
251	CAAACTGTGT GGGCCGCCAG ATTCTTGCGG CCTTGGGACAG CCTGTTgggg
301	cggatcccggtt TTTGTCATACT CATCTATTCG AGTTGAAAAA AAATATCCGA
351	ATGCTCTGCTT TCCGACAGCA GCGCTCTGTT TAAAACCGGC GTACTCGTC
401	CCTGGTCCCCA ATCGGGTATT TGGACCATGT CATTGCGTGC CGGTCAAGTG
451	TGGAAATGGGG TAAAGGGCGC ATTGGCGGAG GATGGCGATT ATCTTTCCGT
501	GTATGCTCCG ACCACGCCCA ACCGGCCCGG CGGTACTAT ATATGGTAA
551	AGAAAAGCGA TGTCGGCGAA CTGATATGA GCGTGGACGA AGCGTTGAAA
601	TATGTGATTG CGCTGGTGTAT GGTCACTCCG GACGACCTGC CGCTCAAAAC
651	ATTGGCAGGG CCTATGGCGC CTGAAAGGG GGAGGTGCC GAACACAAAT
701	AA

This corresponds to the amino acid sequence <SEQ ID 748; ORF98ng-1>:

1	MTEPAAEKGK AAKALKYKL TGLWLPPIA VTVWWVSYIV SASDQLVNLL
51	PKQWRQYVL GNFNIPGLGV VAIAVLFVTG LFAANVLGRQ IILAANDSLLG
101	RIPVVKSIYS SVVKVSESSL SDSSRSFKTP VLVPFPQSGI WTIAFVSGOV
151	SNAVKAAALPQ DGDYLSYYVP TTNPNTGGY IMVKKSVDRE LDMSVDEALK
201	YVISLGMVIP DDLPVKTLAG PMPPEKAELPEQQ*

ORF98ng-1 and ORF98-1 show 97.9% identity in 233 aa overlap:

		10	20	30	40	50	60	
50	orf98-1.pep	MTEXAAEKGKAAKALKYLITGILWLPPIAVTVWWVSYIVSASDQLVNLLPKQWRQYVL						
	orf98ng-1	MTEPRAEKGKAAKALKYLITGILWLPPIAVTVWWVSYIVSASDQLVNLLPKQWRQYVL						
55	orf98-1.pep	10 20 30 40 50 60						
	orf98-1.pep	70 80 90 100 110 120						
		GFNIPGLGVIVIAVLFVTGLFAANVLGRQIILAANDSLLGRIPVVKSIYSSVKVSESSL						

	orf98ng-1	GPNIPGLGVIVAIAVLFVFTGLFAANVLGRQILAAWDSLGLRIPVVKSIVSSVKKVSESLL 70 80 90 100 110 120
5	orf98-1.pep	SDSSRSFEXTPLVLPFPQPGINTIAFVSGQVSNAVKAALEPKQGDYLSVYVPTTNPTGGYY 130 140 150 160 170 180
	orf98ng-1	SDSSRSFEXTPLVLPFPQPGINTIAFVSGQVSNAVKAALEPKQGDYLSVYVPTTNPTGGYY 130 140 150 160 170 180
10	orf98-1.pep	IMVKKSDVRELDMSVDEALKYVISLGWVIPDDLPVXTLAGPMPEKADLPEQQX 190 200 210 220 230
	orf98ng-1	IMVKKSDVRELDMSVDEALKYVISLGWVIPDDLPVXTLAGPMPEKADLPEQQX 190 200 210 220 230

15 Based on this analysis, including the fact that the putative transmembrane domains in the gonococcal protein are identical to the sequences in the meningococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 89

20 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 749>:

```

1 ATgAAAAACGG TAGTCTGGT TGTGCTCTG TTTGCCCGG CCGTCGGACT
51 GCGCTGGCT TCAGGCATT ACACGGGCGA CGTGATATAC GTACTCGGAC
101 AGACATCGT CAGAACAAAC CTGCAACGGCT TTGTTGTTAGG TTGCTGTGATT
151 GCGCTGTGG TGTGTTATT TTGTTAAA TTCAATTATC GgGgTACTCA
201 ATATCCCCGA AAAGATGCG AGTGTGGTT CGCGGCGCTAA AGGCCkCAAG
251 ssCGGGCTTC CCTTGACACA CGGGGGTTTG CGGTATTG AAGGGGTTT
301 TGAAAGGCC GAAACTAGAG CCTCACCGGGT GTGTCGCAAC AAAGLAGGCC
351 GaGAGAACAC CGGACTTTC CATTGATGCT GGGGGCGCAC GCGGCCGAC
401 AGATGGAAAAA CATCGGAACT CGCGGCGCTT ATCTTGCCTA AATGCCCAA
451 CTGCCGAAAC AACAGCGAGT TTCCGGTTAT CTTTGTTG CGGAATCGGC
501 GTTGGACCGG CGGGCATTAG AGGCCGGCGA AGCCAAATCTT CATGCCGGGG
551 CGAGATGTA TGCCPAACTTG ACCGGCTCTG TGCGCTCTGA .ATTCTTAC
601 GCTTTCGACA GGCGGAGACG GTTGCAGGTT CTGCGAAAAA CGGAAAACCT
651 TTGAAAGGCC GCGCGCTTG GCATAATCGA ATUGAACCGG TTCAAAATT
701 GGGCATATTC GTCGCCGAGCT GGGGATGCT GCGGATGCC CGCGCTTGA
751 AACCTGGCTG AAGGGATTTC CGCAGACGGCT CAAAAGCGG GAATTGAGCG
801 TATCGGTTGC GGGAAAGATC GAACTTGG GACTGTATGC CGATCGGTC
851 AAATGGCTCA AACAGCATTC TGCGCasAAC CGCCGGCCCG AGCTTTTGGA
901 AGCGCTTGTG GAAAGCGCTC GCTTTTGGG CGACGGCGA CAGCAGAAAG
40 1001 CCTCGATTTC TCCGGCTGGT TGCGTGAAG AACAGCGCCAA TAACCGGGTT
1051 GCGAAAGGC TACCTTGAG CGAGCATTTG ATTAAAGCGG ATGATTTCGG
1101 CGGGTTTGCT TCTAACAAAG GTTTTCGACG AAATCGGAGA ACCGCAAGAG
1151 GCGGAGGCGC AC...

```

45 This corresponds to the amino acid sequence <SEQ ID 750; ORF100>:

```

1 MKTVVIVVL FRAAVGLALA SGIVYGDVYI VLQQTMLRIN LHFVULGSLI
51 AVVWYFLFK FIIGVLNIE KMQRFGSARK GKXXLALNK AGLAYFEGRF
101 EKAELAERSRV LVNKGVRDNR TLAIMLXAHAA AQQMENIXXR DRYLAEIAKL
151 PERQQLSRYL LLAESALNRR DYEEAAEANLH AAAMMNANLT RLVRLXIRYA
201 FDGRQALQVL AKTEKLSKAG ALGKSEMER YQNWAYRQLA DAADAALKT
251 CLKRIPDLSL NGEELSVAE KYERLGLYAD AVKVKWQHYE XNRRFELLEA
301 FVESTFLGE REQKQAIIDA DAWLKEQPDN ALLMLYGLR AFGRKWLWCKA
351 KGYLEASIAL KPSISARLVL TKVFDIEGP QKAEAH...

```

Further work revealed the complete nucleotide sequence <SEQ ID 751>:

```

55 1 ATGAAACGG TAGTCTGGT TGTGCTCTG TTTGCCCGG CCGTCGGACT
51 GCGCTGGCT TCAGGCATT ACACGGGCGA CGTGATATAC GTACTCGGAC
101 AGACATCGT CAGAACAAAC CTGCAACGGCT TTGTTGTTAGG TTGCTGTGATT
151 GCGCTGTGG TGTGTTATT TTGTTAAA TTCAATTATC GCGTACTCA

```

201 TATCCCCGAA AAGATGCAGC GTTTCGGTTIC GGCAGCTAAA GGCGCGAAGG
 251 CCGGGCTTGC CTTGACAAAGC GCGGGTTTGG CGTATTITGA AGGGCGTTTT
 301 GAAAGCCGG AACTAGAACG CTACACCGTG TTGGTCACAA AAAGAGGCCGG
 351 AGACAAACCGG ACTTTGCGAT TGATGCTGGG CGCGAGCGCC CGCGGACAGA
 401 TGGAAACAT CGACGCTGCG GACCGTATAC TTGCGGAAAT CCCAAACCTG
 451 CGGAAARAAAC AGCAGCTTTC CGGTATCTT TTGTTGCGG AGATCGCGTT
 501 GAACCGCGC GATTACGAAG CGCGGAAAGT CAATCTTCAAT GGGCGCGGA
 551 AGATGAATGC CAACTTACG CGGCCCTGGT GCTGCGCAACT TGTTACGCT
 601 TTGCGAGGGG GCGAGCGTT GCAGGTCTTC GCAAAACCG AAAACCTTC
 651 CAAGGGGGGG GCGTTGGCGA AATCGGAAAAGG GAAACGGTAT CAAAATTGGG
 701 CATACCGCGC CCAGCTGGCG GATGCTGCC ATGCCGCCG TTTGAAAACC
 751 TGCCGTAAGG GGATTCGGCA CAGCTCTCAA AACCGGGAAAT TGAGCGCTATC
 801 GGGTGGGAA AAGTACGAAC TTGTTGGACT GTATGCCGAT CGGGTCAAAT
 851 GGGTCAACAG CATTATATCG CACAACCCGC GCGCCGGAGCT TTGCGAGGC
 901 TTGTCGAAAC CGGTGCGCTT TTGGGGCGAG CGCGAACAGC AGAACGCAT
 951 CGATTTGCGC GATGCTGGC AGTAAAGAACAA GCGCGATAAC CGCGTTCTGC
 1001 TGATGTTACTC CGGTGCGCTC GCCTACGGCC GCAAACTTIG GGGCAAGGCA
 1051 AAAGGCTTACG TTGAGGCGAG CATTGCGATTA AAAGCGGAGTA TTTCGCGGG
 1101 TTGTTGCTCA GCAAAAGTTT TCAGCACAAAT CGGAGAACCG CAGAAAGCGG
 1151 AGGGCAGGG CAACTGTGTT TTGGAAGCGC TCTCGATGCA CGAACGTCAC
 1201 GCAGCGTTAG AGCACATAG CTGA

This corresponds to the amino acid sequence <SEQ ID 752; ORF100-1>:

1 MKTVVWIVVL FAAAWGLALA SGIVTDVYI VLQGTMRLIN LHAFVLSLII
 51 AVVVVWFLEK FTIGVLNIE KMQRFGSARK GRKAALALIN AGLAYEFGRF
 101 EKAEASRSV LVNKEAGDNR TLALMLGHAQ ACQMMENIELR DRYLAEIAKL
 151 PEKQSLRSYL LLAESALNR DYEAEANLH AAAKMNANLT RLVRLQLRYA
 201 FDGRDALQVL ARTKEKLSKAG ALGKSEMERY QNWAYRQLA DAADAAALKT
 251 CLRKIPDSLXK NGELSVSVAE KYERLQLYAD AVKWKOHYP INRRPELLEA
 301 FVEVSFRLE REQPKAIDFA DAWLKEQPDN ALLIMYLGLR AYGRKLWGKA
 351 KGYLEASIAL KPSISARLVL AKVFDEIGEP QKAEAQRNIV LEAVSDDERH
 401 AALEQHS*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF100 shows 93.5% identity over a 386aa overlap with an ORF (ORF100a) from strain A of *N.*

meningitidis:

	10	20	30	40	50	60
orf100.pep	MKTVVWIVVL FEA 	AAGVLA 	LASGIY 	TGDVYI 	VLQGTMRL 	IN LHAFVLSLII
orf100a	MKTVVWIVVL FEA 	AAXGLA 	LASGIY 	TGDVYI 	VLQGTMRL 	IN LHAFVLSLII
	10	20	30	40	50	60
orf100.pep						
orf100a						
	70	80	90	100	110	120
orf100.pep	FIIGVLNIP 	PEKMQRFGSARK 	GXXXLANKA 	GLAYFE 	CGREKAE 	LEASRV
orf100a	FIIGVLNXP 	PEKMQRFGSARK 	GRKAALAL 	AGLAYFE 	CGREKAE 	LEASRV
	70	80	90	100	110	120
orf100.pep						
orf100a						
	130	140	150	160	170	180
orf100.pep	TLALMLXAH 	AAQGMENIX 	RDRYLAEIAKL 	PEKQQLSRY 	LLESALNRR 	DYEAAEANLH
orf100a	TLALMLXAH 	AAQGMENIEL 	RDRYLAEIAKL 	PEKQQLSRY 	LLESALNRR 	DYEAAEANLH
	130	140	150	160	170	180
orf100.pep						
orf100a						
	190	200	210	220	230	240
orf100.pep	AAAKMNANL 	TRLVRXLIRYAF 	DRGDAQLV 	LAKTEKL 	SKAGALGK 	SEMERYQN
orf100a	AAAKMNANL 	TRLVRXLIRYAF 	DRGDAQLV 	LAKTEKL 	SKAGALGK 	SEMERYQN
	190	200	210	220	230	240
orf100.pep						
orf100a						
	250	260	270	280	290	300
orf100.pep	DAADAAALKT 	CLKRIPDSL 	LNKGELS 	VSVAAE 	KYERLGLY 	DAVVKWKOHYP
orf100a	DAADAAALKT 	CLKRIPDSL 	LNKGELS 	VSVAAE 	KYERLGLY 	DAVVKWKOHYP
	250	260	270	280	290	300

-421-

		310	320	330	340	350	360
	orf100.pep	FVESVRLFLGEREQKAIADFADAWLKEQPDNALLMLYGLRLAGRKLNGKAKGYLEASIAL					
5	orf100a	FVESVRLFLGERDQQKAIADFADAWLKEQPDNALLXYLGLRLAYGRKLNGKAKGYLEASIAL	310	320	330	340	350
							360
		370	380				
10	orf100.pep	KPSISARLVLTKVFDEIGEPQKAEEAH					
		::: : : :					
	orf100a	KPSISARLVLAKVFDETGEPEQKAEQARNLVLASVAEENRPSAETH	370	380	390	400	

The complete length ORF100a nucleotide sequence <SEQ ID 753> is:

15	1	ATGAAAACGG TAGTCGCGAT TGTGCGCCG CKNNTCGGGCT					
	51	GGCATCTGGCG TCGGCGATTC ACACCGCGA CCTGTATATC GTACTCGGAC					
	101	AGACCATGCT CGAGATCAAC CTGCAAGCGCT TTGTTGTAAG TTGCGCTGATT					
	151	CGCGCTGCG TCTGGTATTCT CCTGTTCAAAT TCATCATCG GCGACTCTAA					
	201	TANCCCGGAA AAGATGCAAG GTTTCGGGTC GGCGCTTAAGA GCGCGCAAGG					
	251	CCGGCTGCTG TTGACACAGC GCGGGTTTGC CTATTTTGAG AGGCCCTTTT					
20	301	GAAAAGCGG ATACTTGAAGC CTGGCGCGTA TTGGGAAACA AAGAGGGGGG					
	351	GGATAACCGG ACTTCTGGCAT TGATGTTGGC CGCACATGCCC GCGGGCGAGA					
	401	TGAAACACAT CGAGCTCGC GACCGTATTC TTGCGGAAT CGCAAACATC					
	451	CCGGAAGAAC AGCACCTTC CGGTATATCT TTGTTGGCGG ATATGGCGTT					
	501	GAACCGGGG GATTACGAAG CGCGCGGAAGT CAATCTTCAT CGCGCGCGA					
25	551	AGATGATGTC CAACTTAGG CGCTCGTGC GTCTGCACACT TGCTTACGCT					
	601	TTTCGACAGGG CGCGACGGTT GCAGGTCTTC GCAAAACCG AAAAANTTC					
	651	CAAGGGGGGG GCGTNGGGCA ATCTGGAAAGA GGAAAGCTTAT CAAAATTTGG					
	701	CATACCGGCC CCAGCTGNCG GATGTCGCC ATGCCGCGC TTTGAAAACC					
	751	TGCGCTGAAGC GGATTCCTCGA CAGCTCTCAA AAACGGGGAAAT TGAGCGTATC					
30	801	GGTTGGGAAAG ATGACCAAC GATTTGGACT GTATCGCGAT CGGGTCAAT					
	851	GGGTCAACAA CGATTATCGC CAAACCGGC GACCCGAACAT TTGGAAGGCN					
	901	TTTGTGCAAAG CGCTGCGCTT TTGGGCGAA CGCGATCAGC AGAAAGCAT					
	951	CGATTTGCGC GTATGTTGGC TGAAGAACACA GCGCGATAAT CGCTTCTGC					
	1001	TCGATGATCT CGCTCGCGTC GCCTTACCGCC GCAAATCTTG CGCGAAGCA					
35	1051	AAAGGCTTACG TTGAACGAG CATTGCAATTA AAGCCGAGTA TTTCGCGCG					
	1101	TTTGTGCTG CCAAATGGTT TTGACGAAAC CGGAGAACCG CAGAGGGCG					
	1151	AGGGCAGGG CAACTTGGTT TTGCGCAAGGG TTGCGCGAGGA AAACCGNCCT					
	1201	TCCCGCGAAAC CCCATTGA					

This encodes a protein having amino acid sequence <SEQ ID 754>:

40	1	MKTIVWIVVFAAAXGLALAGIIXTGDVYIVLGQTMLRINLHAFVLGSILIAVVVWYFLK					
	51	AVVWVYFLFL FIIGVLNKEPQMRFGSARKGRKAALALNKAFLAGLAYFEGRFKAELEASRVLGNKEAGDR					
	101	EKAELEASRV LGNKEAGDRN TLALMLGAHAG QAGOMENIELR DRYLAEIAKL					
	151	PEKQQLSRYL LLAESALNRR DYEEAEANLIA AAKRNQNLIT RLVRQLRLYA					
	201	FDRGDAOLVI AKTEKXSKAG AXKGSEMERY QNWYARRQLX DAADAAAALKT					
45	251	CLKRPDSLX NGELSVLSVAE KVERLGLYAD AVKWKVKGHYE HNRPEPLEA					
	301	FVESVRLFLG RDKQKALDF ADWLKQPDN ALLLXYLGLR AYGRKLWKGKA					
	351	KGYLEASIAL KPSISARLVL ARKVFDETGEF QKAEQARNLV LASVAEENR					
	401	SAETH*					

ORF100a and ORF100-1 show 95.1% identity in 406 aa overlap:

50	orf100a.pep	10	20	30	40	50	60
		MKTIVWIVVFAAAXGLALAGIIXTGDVYIVLGQTMLRINLHAFVLGSILIAVVVWYFLK					
	orf100-1	MKTIVWIVVFAAAXGLALAGIIXTGDVYIVLGQTMLRINLHAFVLGSILIAVVVWYFLK	10	20	30	40	50
							60
55		70	80	90	100	110	120
	orf100a.pep	FIIGVLNKEPQMRFGSARKGRKAALALNKAFLAGLAYFEGRFKAELEASRVLGNKEAGDR					
	orf100-1	FIIGVLNKEPQMRFGSARKGRKAALALNKAFLAGLAYFEGRFKAELEASRVLGNKEAGDR	70	80	90	100	110
							120
60	orf100a.pep	130	140	150	160	170	180
		TLALMLGAHAAGQMENIELRDRYLAIEIAKLPEKQQLSRYLLAESALNRRDYEEAEANLH					
	orf100-1	TLALMLGAHAAGQMENIELRDRYLAIEIAKLPEKQQLSRYLLAESALNRRDYEEAEANLH	130	140	150	160	170

-422-

		130	140	150	160	170	180
5	orf100a.pep	190	200	210	220	230	240
	orf100-1	AAAKNNANLTRLVRLQI.RYAFDRGDAIQLVLA KTEKXS.KAGAXGKSE.MERYQNWAYRQLX 190	200	210	220	230	240
10	orf100a.pep	250	260	270	280	290	300
	orf100-1	DAADAAALKTCLKRIPDSIKNGE.LSVVAE KYERLGLYADAVKWKQHYPHNRRPELLEA 250	260	270	280	290	300
15	orf100a.pep	310	320	330	340	350	360
	orf100-1	FVE.SVRF.LGERD.QQKAID.FADAWLKE.QP DNALLXYL.GRLAYGRKLWGKAKGYLEASIAL 310	320	330	340	350	360
20	orf100a.pep	370	380	390	400		
	orf100-1	KPSISARLVLAKVFD.EIGEP.QKAB.AQRNL VLEAVSDDERHAALEQHSX 370	380	390	400		
25							

Homology with a predicted ORF from *N.gonorrhoeae*

ORF100 shows 93.3% identity over a 386 aa overlap with a predicted ORF (ORF100ng) from *N.gonorrhoeae*:

30	orf100.pep	MKTVVWIVVL.FAAA.VGLALAS.GIYTGDVY.IVL.GQTMLR.I.NL.HAFVL.GSLIAVVVWF.LFK	60
	orf100ng	METTVVWIVVL.FAAA.VGLALAS.GIYTGDVY.IVL.GQTMLR.I.NL.HAFVL.GSLIAVVVWF.LFK	60
35	orf100.pep	FII.GVLN.IPEK.MQR.FSGARK.GXX.LA.I.NKAGI.LAY.FEGR.FEKAE.LEASRV.LVN.KVG.RDNR	120
	orf100ng	FII.GVLN.IPEK.MQR.FSGARK.GKR.KA.I.LNKAGI.LAY.FEGR.FEKAE.LEASRV.LVN.KG.NDNR	120
40	orf100.pep	TIA.I.MLXA.HAA.QMEN.I.NX.RD.YL.A.T.I.A.KL.PE.Q.QLS.RY.LI.I.AESAL.NR.DYE.AE.A.N.I.H	180
	orf100ng	TIA.I.MLGA.HAA.QMEN.I.E.RD.YL.A.T.I.A.KL.PE.Q.QLS.RY.LI.I.AESAL.NR.DYE.AE.A.N.I.H	180
45	orf100.pep	AAAK.MN.NLTRL.VRLX.I.RYAF.DRGDAI.QVLAKTE.KLS.KAGAL.GKSE.MERYQN.WAYR.QLA	240
	orf100ng	AAAK.MN.NLTRL.VRLQI.RYAF.DRGDAI.QVLAKTE.KLS.KAGAL.GKSE.MERYQN.WAYR.QMA	240
50	orf100.pep	DAADAAALKTCLKRIPDSIKNGE.LSVVAE KYERLGLYADAVKWKQHYPHNRRPELLEA	300
	orf100ng	DAADAAALKTCLKRIPDSIKNGE.LSVVAE KYERLGLYADAVKWKQHYPHNRRPELLEA	300
55	orf100.pep	FVE.SVRF.LGERE.QQKAID.FADAWLKE.QP DNALLMYL.GRLAYGRKLWGKAKGYLEASIAL	360
	orf100ng	FVE.SVRF.LGERE.QQKAID.FADAWLKE.QP DNALLMYL.GRLAYGRKLWGKAKGYLEASIAL	360
55	orf100.pep	KPSISARLVLTKVFD.EIGEP.QKAB.AQRNL VLEAVSDDERHAALEQHSX	386
	orf100ng	KPSISARLVLTKVFD.EIGEP.QKAB.AQRNL VLEAVSDDERHAALEQHSX	405

The complete length ORF100ng nucleotide sequence <SEQ ID 755> is:

60	1	ATGAAAACGG TAGTCTGGAT TGTGTCCTG TTTCCCGCGC CCGTCGGACT	
	51	GGGGCTGGCT CGGGCATTT ACAGCCGGC CGTGTTATATC GTACTCGGAC	
	101	AGACATGCTC CAGATCACR CTGCACGGCT TTGTTGTTAG TTGCTGTATT	
	151	GCGGTCTGG TTGCGTATTCT CTGTTGTTAAA TTCTCATCG GCTATACTCA	
	201	TATCCCGGAA AATATGGGGC GTTCGGGTTC GGCGCCGAAA GGCGCCGAAGG	
	251	CCGGCTGGTCT CTGTAATAGG CGGGTTGG CGTATTTCCGA AGGGCGTTTT	
	301	AAAAAGGGGG AACCTGAAGC CTCTCGAGTG TTGGGCAACA AAGAGGGCGG	
	351	AGCAACCGG ACTTCTGGCAT TGATGCTGGG CGCCACGG GCAGGACAGA	
65	401	TGGAAATAAT CGAGCTGGCG GACCTTATC TTGCGGAAT CGCCAACCTG	

451 CGGGAAAAAC AGCACGTTTC CGCGTATCCTT CTGCTGGCGG AATCGGGCTT
 501 AAACCGGG GATTAGGAAG CGGGCGGAAC CAACTTCTAT GCGGGCGGGCA
 551 AGATGAATGC CAACCTTAAG CGGCCGTGTC GTCTGCAACT TCCTTACGCC
 601 TTCCATCGGG CGCATGGCTT GCAGGTTCTG CCAAAAaccG AAAAATTTC
 5 CARGGCGGG CGCTTGGGCA AATCGGAATM GGAACGGTAT CAAATTGGG
 701 CATACCGCC CGAGATGGG GATGCTGCCG ATGCCGCCG TTGAAAACC
 751 TGCCGTAAAGC GGATTCGGG CAGCCTCAA AACCGGGAAAT TGAGcGTATC
 801 GGTTGCGGAA AGATACGAA GTTGGGACAT GTATGCGGAT GCGGTCAAAT
 851 GGTCACAAAC GCATTTATCG CACAACCGG GCCCGGAGCT TTGGAGGCC
 901 TTTCGCAAA GCGTGGCCTT TTGGGGAG GCGGAACAGC AGAAAGCCAT
 951 CGATTGTCG GATTCTTGGC TGAAGAACAC GCGCGATAAC GCGCTTCTGC
 1001 TGATGTATCT CGGGCGGCTC GCCTACGCCG CCAAACTTGG GGGTAAGGCA
 1051 AAAGGCTACC TTGAGGCGG TATTCGACTG AAGCCAGTA TTCCGGCCG
 1101 TTGGGTGTTG GCAAGGGTTT TTGACGAAAC CGCACAGTCG CAAAAGCCG
 1151 AACGACACGCC CAACTTGTT TTGCGCAAGG TTGGCGGGGA AAACCGCCCT
 1201 TCCGGCGAA CCCCTTGA

This encodes a protein having amino acid sequence <SEQ ID 756>:

1 MKTVVWIVVL FAAAVGLALA SGIVYGDVYI VLGQTMRLIN LHAFVLGSLI
 20 51 AVVVVWIFLK PIIGVINITV NMRSSGARK GRKAALALNK AGLAYFEGRF
 101 KEAEALASRV LGNKEADGN TLAJMLGHNA AQGMENIELR DRYLAEIAKL
 151 PERKQLSRYI LLAESALNRD DYEEAAEANLH AAAMMNANLIT RLVRQLRHYA
 201 FDGRDALQVLT AKTEKLSKAG ALGKSEMERY QNWAYRRQMA DAADAALAKT
 251 CLKRIPDSLK NGELSVSVAE KYERLGLYAD AVKWKVQHYIP HNRPPELEA
 301 FVEVSRLFLGE REQQKAIDFA DSWLKEQPDN ALLLMLYGLR AYGRKLWCKA
 351 KGYLEASIAL KPSIPARLVL AKVFKDETAQS QKAEPQRNLV LASVAGENRP
 401 SAETR*

ORF100ng and ORF100-1 show 95.3% identity in 402 aa overlap:

	10	20	30	40	50	60
30 orf100-1.pep	MKTVVWIVVLFAA	VGLALA SGIVYGD	VYI VLGQTMRLIN	HAFVLGSLI	A VVVVWIFLK	
orf100ng	MKTVVWIVVLFAA	VGLALA SGIVYGD	VYI VLGQTMRLIN	HAFVLGSLI	A VVVVWIFLK	
	10	20	30	40	50	60
35 orf100-1.pep	70 FIIGVNLNPEK	70 MQRFGSGARKGRKA	80 ALLNKAGLAYFEG	90 GREKAELEASRV	100 VLNVKEAGDNR	110 120
orf100ng	FIIGVNLNPEKMR	RSSGARKGRKA	ALLNKAGLAYFEG	GREKAELEASRV	VLNVKEAGDNR	
	70	80	90	100	110	120
40 orf100-1.pep	130 TLALMLGAHAAG	130 QMENIELRDRYLA	140 ETAKLPKQQLSRY	150 LLAE SALNRRDYE	160 EEAANLH	170 180
orf100ng	TLALMLGAHAAGQ	MENIELRDRYLA	ETAKLPKQQLSRY	LLAE SALNRRDYE	EAEANLH	
	130	140	150	160	170	180
45 orf100-1.pep	190 AAAKMNANL	190 TRLVRQLRYA	200 FDRGDAQVLA	210 KTEKLPSKAG	220 ALGKSEMERY	230 QNWAYRRQMA
orf100ng	AAAKMNANL	TRLVRQLRYA	FDRGDAQVLA	KTEKLPSKAG	ALGKSEMERY	QNWAYRRQMA
	190	200	210	220	230	240
55 orf100-1.pep	250 DAADAAALKT	250 CLKRIPDSL	260 KNGELSVSVAE	270 KYERLGLYAD	280 AVKWKQHYIP	290 HNRPPELEA
orf100ng	DAADAAALKT	CLKRIPDSL	KNGELSVSVAE	KYERLGLYAD	AVKWKQHYIP	HNRPPELEA
	250	260	270	280	290	300
60 orf100-1.pep	310 FVEVSRLFLGE	310 REQQKAIDFA	320 DWLKEQPDN	330 ALLLMLYGLR	340 AYGRKLWCKA	350 LEASIAL
orf100ng	FVEVSRLFLGE	REQQKAIDFA	DWLKEQPDN	ALLLMLYGLR	AYGRKLWCKA	LEASIAL
	310	320	330	340	350	360
65 orf100-1.pep	370 KPSISARLVL	370 LAKVFDIE	380 GIEPQKAE	390 QRNLVLEA	400 VSDDERHAA	LEAHLEQHSX
orf100n	KPSISARLVL	LAKVFDIE	GIEPQKAE	QRNLVLEA	VSDDERHAA	LEAHLEQHSX
	370	380	390	400		

370 380 390 400

Based on this analysis, including the presence of a putative leader sequence, a putative transmembrane domain, and a RGD motif, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 90

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 757>

10	1	ATGATGTTTCTTGCGGTCAACGTGTTTCAC	TGTTTTTGTCATCTGCCT
	51	GTTTGCAGGGTGTGGTTTACCG	TGGCCGAGGATTTCGCTCAAT
	101	TGATGTCGCCGGCGGCAAT	ATGGCAGTATGTCGCTTCGGGATCTGGCC
	151	GTGGCGCTGTGACCTTTTA	TGCGCTGGTGTGGCTTCGGGCTGTTG
	201	CGGGCGGCCATACAGGGTTG	CGGGCGGCTGTCGGGGCCAGCGCTGTTG
	251	ACGCTTACCTGTGGTTGGC	TTCGATGTCATTCGCTGTTTACCA
	301	GCGCTGCTGCCTGGCGCTT	TGGCTGTTACACGAATGTTTACCGG
	351	CTGGTAGCTACCGG	TGTTTCAACAACTCCCGTGTGCTGTGATG
	401	TGTATGTTGCTCTGTTCAA	CCCTTGTGATGTTGCGCCGGG

This corresponds to the amino acid sequence <SEQ ID 758; ORF102>;

20 1 MMFSWFKLFH LFFVISWFGA LFYPLPRIFVN MAMIDVPRGN PEYVRLSGMA
51 51 VRLYRFSMEL GFAGAVVEGA IPPFAAGWGWS GWVHVKLCLG LMLLAYQLYC
101 101 GVLLRRFQDY SNASFRHRYWA VENIEPVLLA VAALYXVVFK PF*

Further work revealed the complete nucleotide sequence <SEQ ID 759>;

	1	ATGATGTTTT	CTTGGCTTCA	GCTGTGTTAC	TGTTTTTG	TCTATTGCTG
25	51	CTTTGCAGGG	CTGTTTAAAC	TGCCGAGGAT	TTCCGTCATA	ATGGCATGATC
	101	TTGATGTCG	GGGGCGGCAT	GAAGGATATG	TGGCTGTC	GGGATCATGGC
	151	GTGGCGCTGT	ACGGTTTGT	GTGCGGCTGT	GGCTTGGCC	CCTGTGTTG
	201	CGGGCGGCAG	ATACCGGTTC	CGCCGGCTGT	TGGGGCCAGG	GGCTGTGAC
	251	ACCTCAACG	TGGTTGGG	TTGATGCTCT	TGGCTTACCA	TGGTATGTC
30	301	GGCGCTGCG	TGGCGCGGT	GGGATTTAC	ACGAACTG	TTTACACCCG
	351	CTGGTACCGG	GTGTTCAACG	AAATCCCGGT	GCTGCTGATG	GTGGCCGGCG
	401	TGTATCTGG	CCTGTCATAA	CGCTTGTGA		

This corresponds to the amino acid sequence <SEO ID 760: ORF102-1>:

1 MMFSWFKLFH LFFVISWFGA LFYLPRLIFVN MAMIDVPRGN PEYVRLSGMA
51 VRLYRFMSPL GFCAVGVEGA IPFAAGGWGS GWVHVKLCLG LMLLAYQLYC
101 GVLLRFDYF SNAFHSRWR YVLLVLLM VAALYLVVEK PE*

Computer analysis of this amino acid sequence gave the following results:

Homology with HP1484 hypothetical integral membrane protein of *H. pylori* (accession number AE000647)

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF102 shows 99.3% identity over a 142aa overlap with an ORF (ORF102a) from strain A of *N.meningitidis*:

5	orf102.pep	MMFSWFKLFHLLFVISWFAGLFYLPRLPRIFVNMMAMIDVPRGNPEYVRLSGMAVRLYRFLMSPL	10	20	30	40	50	60
	orf102a	MMFSWFKLFHLLFVISWFAGLFYLPRLPRIFVNMMAMIDVPRGNPEYVRLSGMAVRLYRFLMSPL	10	20	30	40	50	60
10	orf102.pep	GFGAVVFGAAIPFAAGWWGSGWVHVVKLCLGMLLAYQLYCGVLLRRFQDYSNAFSHRWYR	70	80	90	100	110	120
	orf102a	GFGAVVFGAAIPFAAGWWGSGWVHVVKLCLGMLLAYQLYCGVLLRRFQDYSNAFSHRWYR	70	80	90	100	110	120
15	orf102.pep	VFNEIPVLLMVAALYLVVFKPFX	130	140				
	orf102a	VFNEIPVLLMVAALYLVVFKPFX	130	140				

The complete length ORF102a nucleotide sequence <SEQ ID 761> is:

25	1	ATGGATTTCTTGTGGTCAC	20	30	40	50	60
	51	GTTTGCAAGG	CTGTTTTTAC	TGCCAGAGT	TTTCGTCATA	ATGGCGATGA	
	101	TTGATGTGGC	GCGCGCAAT	CCCGAGATAG	TGGCTCTGTC	GGGCATGCGC	
	151	GCGGGCTGT	ACCGTTTAT	GTCGCCCTTG	GGCTTGCGC	CGCTGGTGT	
	201	CGGCGCGCGG	ATACCGTTAG	CGCGCGCTGC	GTGGGGCAGC	GCCTCGGGTAC	
	251	ACGTCAAACT	GTGTTGGGC	TTGATGCTCT	TGGCTTACCA	GTGGTATTGC	
	301	GGCGGTGTC	TGGCGCTTT	TCAGGATCTAC	AGCAATGCTT	TTTACACACCG	
	351	CTGGTACCGC	GTGTTCAACG	AAATCCCGCT	GCTGCTGATG	GTGCGCCGC	
	401	TGTATCTGGT	CGTGTTCAAA				

This encodes a protein having amino acid sequence <SEQ ID 762>:

35	1	MMFSWFKLFH_LFFVISWFAGLFYLPRLPRIFVN_MAMIDVPRGNPEYVRLSGMAVRLYRFLMSPL	20	30	40	50	60
	51	VRLYRFLMSPL	IPFAAGWWG	SGWVHV	VKLCIG	LMLLAYQLYC	
	101	GVLLRRLQDY	SNAFSHRWY	VFNEIPVLLM	VAALYLVVF	KPF*	

ORF102a and ORF102-1 show complete identity in 142 aa overlap:

40	orf102a.pep	MMFSWFKLFHLLFVISWFAGLFYLPRLPRIFVNMMAMIDVPRGNPEYVRLSGMAVRLYRFLMSPL	10	20	30	40	50	60
	orf102-1	MMFSWFKLFHLLFVISWFAGLFYLPRLPRIFVNMMAMIDVPRGNPEYVRLSGMAVRLYRFLMSPL	10	20	30	40	50	60
45	orf102a.pep	GFGAVVFGAAIPFAAGWWGSGWVHVVKLCLGMLLAYQLYCGVLLRRFQDYSNAFSHRWYR	70	80	90	100	110	120
	orf102-1	GFGAVVFGAAIPFAAGWWGSGWVHVVKLCLGMLLAYQLYCGVLLRRFQDYSNAFSHRWYR	70	80	90	100	110	120
50	orf102a.pep	VFNEIPVLLMVAALYLVVFKPFX	130	140				
	orf102-1	VFNEIPVLLMVAALYLVVFKPFX	130	140				

55 Homology with a predicted ORF from *N.gonorrhoeae*

ORF102 shows 97.9% identity over a 142 aa overlap with a predicted ORF (ORF102ng) from *N.gonorrhoeae*:

-426-

	orf102.pep	MMFSWFKLFHLFFVVISWFAGLFYLPRIFVNMMAMIDVPRGNPEYVRLSGMAVRLYRFLMSPL	60
	orf102ng	MMFSWFKLFHLFFVVISWFAGLFYLPRIFVNMMAMIDVPRGNPEYVRLSGMAVRLYRFLMSPL	60
5	orf102.pep	GFGAVVFGAAIPFRAAGWWSGWWVKLCLGLMLLLAYOLYCGVLLRRFQDYSNAFSHWRWYR	120
	orf102ng	GFGAVVFGAAIPFRAAGWWSGWWVKLCLGLMLLLAYOLYCGVLLRRFQDYSNAFSHWRWYR	120
10	orf102.pep	VFNEIPVLLMVAALYLVVFKPF	142
	orf102ng	VFNEIPVLLMVAALYLVVFKPF	142

The complete length ORF102ng nucleotide sequence <SEQ ID 763> is:

1	ATGATGTTT	CTTGGTTCAA	GCTGTTTCAC	TTGTTTTTG	TCAATTCTGT
5	51	GTTTCGAGGG	CTTGTGTTAC	TGCCGAGGAT	TTTCGCTCAAT
15	101	TTCATGCGGC	CCGGGGCCTAAT	CCGGAGATCT	ATGGCGATGTA
20	151	GTGCGGTTGT	ACCGCTTAT	GTGCGCTTTC	TGCGCTCTGTC
25	201	CGCCCGGGCC	ATACCCGTTG	CGCCGCTTGC	GGGGATGGCG
30	251	ACGCTTACACT	GTGTTTGGGC	TTGACGCTCT	TGCGCTATCA
35	301	GGCGCTGCGC	TGCCCGTTT	TCAGGATTTAC	ACGAAATGCTT
40	351	CTGGCTACCCG	GTGTTTCAcg	AATATCCCGT	TTCACACCG
	401	TGTATCTGCT	CCTGTTCAA	CGCTTTTA	

This encodes a protein having amino acid sequence <SEQ ID 764>:

1	MMFSWFKLFH	LFFVISWFAG	LFYLPRI	VN MAMIDAPRGN	PEYVRLSGMA
5	51	VRLYRFMSPL	GFGAVVFGAA	IPIFAAGRWGS	GWVHVKLCLG
25	101	GVLRLRFQDY	SNAFSHWRWYR	VFNEIPVLLM	VAALYLVVFK PF*

ORF102ng and ORF102-1 show 98.6% identity in 142 aa overlap:

	10	20	30	40	50	60
30	orf102-1.pep	MMFSWFKLFHLEFFVISWFAGLFYLPRIFVNMMAMIDVPRGNPEYVRLSGMAVRLYRFLMSPL				
	orf102ng	MMFSWFKLFHLEFFVISWFAGLFYLPRIFVNMMAMIDAPRGNPEYVRLSGMAVRLYRFLMSPL	10	20	30	40
			10	20	30	40
35	orf102-1.pep	GFGAVVFGAAIPFRAAGWWSGWWVKLCLGLMLLLAYOLYCGVLLRRFQDYSNAFSHWRWYR	70	80	90	100
	orf102ng	GFGAVVFGAAIPFRAAGWWSGWWVKLCLGLMLLLAYOLYCGVLLRRFQDYSNAFSHWRWYR	70	80	90	100
			70	80	90	100
40	orf102-1.pep	VFNEIPVLLMVAALYLVVFKPF	130	140		
	orf102ng	VFNEIPVLLMVAALYLVVFKPF	130	140		

In addition, ORF102ng shows significant homology to a membrane protein from *H.pylori*:

45	gi 2314656 (AE000647) conserved hypothetical integral membrane protein [Helicobacter pylori] Length = 148 Score = 79.2 bits (192), Expect = 1e-14 Identities = 50/147 (34%), Positives = 68/147 (46%), Gaps = 13/147 (8%)
50	Query: 3 FSWKLFHLFFVISWFAGLFYLPRIFVNMMAMIDAPRGNPEYVRLSGMAVRLYRFLMSPLGF 62 F W K FH+ VISW A LFYLPRI+FV A + V++ +LY F++ Sbjct: 8 FLWVKAFHVIAVISWMAFLYPLPRLFVYHAENAHKEFVGVVQIQEKF--KLYSPFIASPM 65
55	Query: 63 GAVVFGAAIP-----FRAAGRWWSGWWVKLCLGLMLLLAYOLYCGVLLRRFQDYSNAFS 115 G + F + G + GW+H KL L ++LLAY YC + R + Sbjct: 66 GETLITGILMLLIEPTLFKSG---GWLHAKLALVULLLAYHFYCKKCMRLELKDPTRRN 121
60	Query: 116 HRWYRVFNEIPXXXXXXXXXXKFPF 142 R+YRVFNE P KPF Sbjct: 122 ARFYRVFNEAPTILMILIVLIVVVKF 148

Based on this analysis, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 91

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 765>:

```

5      1 ATGGCAAAAA TGATGAAATG GGCGCTTGT GCGCCGGTGC CGCCGGCACG
51     51 GGTTGGGC GGATGGCTT AACTGAAGCC CGAGCGCAC GTGCTTGATA
101    101 TTACGGAAAC GGTCAAGCGC GGC // .....
101    101 ATTTCGTTTA CGATTTTGCG CGAACCGGAT ACGCCGATTAA AGGCAGAAGCT
101    101 CGACAGCTG CACCCCGGG TACGCCAGAT GTCCGCTGGGC GGTTACAACAA
101    101 CGAGTAGCGA TACGGCTCTCA ATAGGGCTCT ACTATTATGC CGCTTCTGTT
151    151 GTGCGGAATC CGACGGCTCA ACTCGCCACO GGAGTGAACGA CGCAGAAATAC
201    201 GGTTGAATTC GACCGCGCTGA AAAATGTGCT GATTATTCCG TCCTGTACCG
251    251 TGAAAATTCG CGCCGGCAAG GCCTTTGTGC GGCTGTTGGG TGCGGACCGC
301    301 AAGGGCGGG AACCGGAATC CGCGAACCGT ATGAGAGACA GTATGAAATAC
351    351 CGAACTAAAG AGCCGGTTGA AAGAGGGGGA CAAAGTGTGC ATCTCCGAAA
401    401 TAACCGCCGC CGAGCAACAG GAAAGCGGG AACCGGCCCT AGGGGGCCCG
451    451 CGCGCCGAT AA

```

This corresponds to the amino acid sequence <SEQ ID 766; ORF85>:

```

20      1 MAMKKWAAV AAVAAAAAWG GWS.LKPEPH VLDDITEVRR G .....
51     51 ..... .....
101    101 ..... .....
151    151 ..... .....
201    201 ..... .....
251    251 PIKALLDSDV PGLTTMSSGG YNSSTIDTASH AVYYYARSFV FNPDGKLATG
301    301 MTTQNTVEID GVKVNLIPIS LTVKNRGKA FVRVLGADGK AAEREIRTM
351    351 RDSMNTEVKS GLKEGDKVVI SEITAEEQQE SGERALGGP RR*

```

Further work revealed the further partial nucleotide sequence <SEQ ID 767>:

```

30      1 ..GTATCGGTGG CGCGCGAGGC ATCGGGGCAG ATTAAAGTAC TTATATGCAA
51     51 ACTCGGGCAA CAGGTAAAGAA AGGCGGATT GATTGCGGAA ATCAATTGGA
101    101 CCTCGCAGAT CRATACCTCTC AATACCGAAA AATCCAGATT GGAAACGTTAT
151    151 CAGGGCGAAAG TGGTGTGGC ACAGATTGCA TTGGCGAGCG CGGAGAGAAA
201    201 ATATAAGCGT CGCGGGCGCT TATGGAAAGGA AACCGCCACT CTCAAAGAGG
251    251 ATTGGAAAGG CGCGAACGAGT GCCTTGGCGCA CGGCCAAAGGC CAATGTTGCC
301    301 GACCTGAAGG CTTTATACAG ACAGACAAA ATTTCATCA ATACCGCCGA
351    351 GTCCGGAATTG GCCTTAACCGC GCATTACCGG AACGATSGACG GCACACGGTTG
401    401 TGCGCGATTCT CTCGGAAAGAG GGGCGACACTG TGAAGGGCGC GCAGCTCTAG
451    451 CGGACGATGG TCCAATGGC GAATCTGGAT ATGATGTTGA ACAAAATGCA
501    501 GATTGGCGGA CGGCGATATT CCAAGTGGAA GCGGGGGCG GATATTCTGT
551    551 TTACGATTTT GTCCGAACCG GATACCCCGA TTAAAGGGGA GCTCGACAGC
40      601 GTCGACCCCCG GGGTGTGACG GATCTGTGCG GCGCGTTTACA ACACAGTAC
651    651 GGATGACGGCT TCCANTGGC TCTACTATT TGCGCGTTCTC TTGTGCGGA
701    701 ATCCGGACGG CAAACTGGC ACGGGGATGA CGACCCGAAAGA TACCGTTGAA
751    751 ATCGACGGCG TGAAAAATGT GCTGATTATT CGTGTGCTGA CGTGTGAAAAA
801    801 TCGGGCGCGG AAGGGTGTG TGCGCGTGTG GGGTGCGGAG CGCAAGGGCG
851    851 CGGAGCGGG ATACCGGACC GTATGAGAG ACAGTATGAA TACCGAGATA
901    901 AAAAGGGGGT TGAAGAGGGG GGACAAAGTG GTCATCTCCG AAATAACCGC
951    951 CGCGGAGCAT CAGGAAGCGC CGAACGGCG CCTAGGGCGC CGCCGGCGCC
1001   1001 GATAA

```

This corresponds to the amino acid sequence <SEQ ID 768; ORF85-1>:

```

50      1 ..VSVAQASQO IKILYVVLGO QVKKGDLIAE INSTSQNTNL NTEKSKLSTY
51     51 QAKLVSQAQIA LGSAEKKYKR QAAALNKENAT SKEDLESAQD AFAAAANVANA
101    101 ELKALIROSK ISINTAESL GYTIRITADM GTVVAILEVE QGTVNAQOST
151    151 PTIVOLANLD MMLNKMOIAR DGIITVKVACO DISFTLSB DTPIKAKLDS
201    201 VDPGLTTMSS GGYNNSTDTA SNAVYYARS FVPNPDGKLA TGMTTQNTVE
251    251 VDGPKVNLII PSLTVKNRGK GFKVRVLGAD GKAAEREIRT GMRDSTMNTVE
301    301 KSGLKEGDKV VISEITAEEQ QEGERALGG PPR*

```

Computer analysis of this amino acid sequence gave the following results:

-428-

Homology with a predicted ORF from *N. meningitidis* (strain A)

ORF85 shows 87.8% identity over a 41aa overlap and 99.3% identity over a 153aa overlap with an ORF (ORF85a) from strain A of *N. meningitidis*:

		10	20	30	40	
5	orf85.pep	MAKNNMKWAAVAAAAGWGWS-LKPEPHVLDTETVRGG : : : : : :				
	orf85a	MAKNNMKWAAVAAAAGWGWSYLKEPQQAYITEVTRGDISRTVSATGEISPSNLVS 10 20 30 40 50 60				
10	orf85.pep		80	90	100
	orf85a	TIVQLANLDMLNKKMQIAEGDITVKKGQDISFTILSEPDTPIKAKLDSVDPGILTMSG 210 220 230 240 250 260				
15	orf85.pep	110 120 130 140 150 160 GYNSSTDASNAVYYYARSFVNPDPGKLATGMMTQNTVEIDGVKNLIIPLSTVKNRGGK : : : : : :				
	orf85a	GYNNSSTDASNAVYYYARSFVNPDPGKLATGMMTQNTVEIDGVKNLIIPLSTVKNRGGK 270 280 290 300 310 320				
20	orf85.pep	170 180 190 200 210 220 AFVRVLGADGKAEREIRTGMRDSMNTVEKSGLKEGDKVVISEITAEEQQESGERALGGP : : : : : :				
	orf85a	AFVRVLGADGKAEREIRTGMRDSMNTVEKSGLKEGDKVVISEITAEEQQESGERALGGP 330 340 350 360 370 380				
25	orf85.pep	230 PRRX 				
	orf85a	PRRX 390				
30						

The complete length ORF85a nucleotide sequence <SEQ ID 769> is:

35	1	ATGGCGAAAC	TGATGAAAG	GGCGGCTTT	GGCGGCCTGC	GGCGGCCGAC	TGCTGTCTTAT
	51	GGTTGGGGC	GGATGGCTT	ATCTGAGCC	GCGGCCGAC	TGCTGTCTTAT	
	101	TTACGGAAAC	GTCAGGGCC	GGCGACATCA	GCGGCCGCG	TGCTGTCTTAT	
	151	GGGGGAGATT	CGCCGTCGA	CCTGGTATCC	CTGGCGCCG	GGACATCGG	
	201	GCAGATTAG	AAACTTTAG	TCAACATCGG	GAACAGCTT	AAAAAGGGCG	
40	251	ATTGATGTC	GGAAATAC	TGACGCTGC	AGACCAATAC	GCTCATCAC	
	301	GGAAATACCA	ATTTGGAAC	GTATCAGGGC	AGACCTGGTG	CGCGACAGAT	
	351	TGCTGGGC	AGCGCCGGAG	AAAATAATTA	GGCTGAGCGG	GGCTGTGGA	
	401	AGGATGATG	GGAGCTTAA	GAAGATTTGG	AAAAGCCCAA	GAATGCCCTT	
	451	GCGCCGCCCA	AGAACCTTA	TGCGGACTG	AGACCTCTAC	TGAGGCTTAC	
	501	CAAAATTTC	ATCAATACCG	CGCAGCTTGA	ATTGGCTTAC	ACGGGCAATT	
45	551	CCGCAGCAT	GGAGCCGGAC	TGTTGGCGCA	TTCTCTGTGA	AGGGGCCGCA	
	601	ACTGTCAGC	CGGGCAGTC	TACGGCCAGC	ATTGCTTCAA	TGGGAAATCT	
	651	GGTATGATG	TTGACAAAAT	TGCAAGTTGC	CGAGGGCGAT	ATTACCAAGG	
	701	TGAAAGCGG	CGAGGATATT	TCTGTTTACG	TTTCTGCTGA	ACCCGATACG	
50	751	CGGATTAAGG	CGGAATCTGA	CGACGGTCAC	CCGGCGCTGA	CCAGGATGTC	
	801	GTGCGCGCG	TACACAGCA	GTACGGATAC	GGCTTCCCAT	GGGCTACTA	
	851	ATTATGCC	TTCTGGTTG	CGGAATCGG	CGGCAACAT	CGCCACGGG	
	901	ATGACGAGCG	AGAAATACG	TGAAATCGAC	GCTGTTAAGA	ATGTCTGAT	
	951	TATTCGGCC	CTGGCGCTGA	AAAATCGGG	CGCGCAGGG	TTTGGCGGC	
55	1001	TGTTGGTGC	AGACGGCC	GGCGGGGAC	GGAAATCTCG	GGACCGTGT	
	1051	AGACAGACAT	TGAAATACCA	AGTAAAGAAC	GCTGTTAAGA	AGGGGGACAA	
	1101	ATGGTGTCTA	CTGGAAATA	CGCGCCGCCA	CGACAGGAA	AGGGCGCAAC	
	1151	GGCCCTTACA	CGGGCCGCCG	CGCCGATATA			

This encodes a protein having amino acid sequence <SEQ ID 770>.

1 MAKKMMWAAV AAVAAAAGW GWSYLKPEPQ AAYITETVR
 51 GEISFLNIVS VQAQASGQKLY KLVKGGLQQV KGGDLIAEIN STSQNTL
 101 EKSLEYTQKA KVLSQAIQLG SAEKKYRKRQ ALWKDQDATA EDLESADQL
 151 AAAAKKAKM TALIKRQSINT ASEAELSGY TRITATMDCT VVALLVEQG
 201 TVNAAGSTPTV IQLVNLDMN LMNRQIAEBG ITKVKACQDI SFTLSEPD
 251 PIKAKLSDV GLPLTTMSLW YNNSSTDASII AVYYYARSDF PNDGKLT
 301 MTTNTTWRD GVKNVLILPS LYTKRNRRGA FTVRLGACDQ AAEREIRTG
 60
 65

351 RDSMNTEVKVS GLKEGDKVVI SEITAAEQQE SGERALGGPP RR*

ORF85a and ORF85-1 show 98.2% identity in 334 aa overlap:

		30	40	50	60	70	80		
5	orf85a.pep	PQAAYITETVRRGDISRTVSATGEISFPNLVSGAQASGQ1KKLYVVLGGQVKRGDLIARE							
	orf85-1			VSGAQASGQ1KKLYVVLGGQVKKGDDLIARE					
10	orf85a.pep	90 100 110 120 130 140							
	orf85-1	INSTSQTNTLNTEKSKLEYAQAKLVSQA1ALGSAEKKYKRQAALWKAEDLESAQD 40							
15	orf85a.pep	150 160 170 180 190 200							
	orf85-1	ALAAAANVAAELKALIRGSKISINTAESELGYTRITATMDGTVVAILVEEGQTVNAROST 100							
20	orf85a.pep	210 220 230 240 250 260							
	orf85-1	PTIVQLANLDMLNKMQIAEGDITVKVAGQDISFTILSEPDPTPIKAKLDSVDPGLTMMSS 160							
25	orf85-1	PTIVQLANLDMLNKMQIAEGDITVKVAGQDISFTILSEPDPTPIKAKLDSVDPGLTMMSS 160							
	orf85a.pep	270 280 290 300 310 320							
30	orf85-1	GGYNNSSTDATASNAYYYYARSEFPNPDKLATGTTQNTVEILDGVKNVLLIPSHTVKNRGG 220							
	orf85a.pep	GGYNNSSTDATASNAYYYYARSEFPNPDKLATGTTQNTVEILDGVKNVLLIPSHTVKNRGG 220							
35	orf85a.pep	330 340 350 360 370 380							
	orf85-1	RAFVRVLGADCKAAEREIRTGMRDMSNTEVKSGLKEGDKVVISITAAEQESGERALGG 280							
40	orf85a.pep	390							
	orf85-1	PPRXX PPRXX							

Figure 19D shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF85a..

Homology with a predicted ORF from *N.gonorrhoeae*

45 ORF85 shows a high degree of identity with a predicted ORF (ORF85ng) from *N.gonorrhoeae*:

ORF85	1 MAKMKWAAVAAVAAAAAWGGWS .LKPEPHVLDITETVRRG	40
ORF85ng	1 MAKMKWAAVAAVAAAAAWGGWSYLKPEPQAAYITEAVRRGDISRTVSAT	50
50	ORF85ISFTILSEPD 250
	ORF85ng	201 TVNAACGSTPTIVQLANLDMLNKMQIAEGDITVKVAGQDISFTILSEPD 250
55	ORF85	251 PIKAKLDSVDPGLTMMSSGGYNSSTDATASNAYYYYARSEFPNPDKLATG 300
	ORF85ng	251 PIKAKLDSVDPGLTMMSSGGYNSSTDATASNAYYYYARSEFPNPDKLATG 300
60	ORF85	301 MTTQNTVEIDGVKNVLLIPSHTVKNRGKAFVRVLGADGKAAEREIRTG 350
	ORF85ng	301 MTTQNTVEIDGVKNVLLIPSHTVKNRGKAFVRVLGADGKAAEREIRTG 350
65	ORF85	152 RDSMNTEVKSGLKEGDKVVISITAAEQESGERALGGPPRR 393
	ORF85ng	351 KDSMNTEVKSGLKEGDKVVISITAAEQESGERALGGPPRR 393

The complete length ORF85ng nucleotide sequence <SEQ ID 771> is:

```

1 ATGGCAAAAA TGATGAAATG GGCGCTGTG CGCGCGGTCG CGGCAGGCaac
51 GGTTGGGGC GGATGGTCTT ATCTGAAGCC CGAACCGCAG GCTGCTTATA
101 TTACGGAAac ggTCAGGCCG GGGCATATCA CGGGGACGGT TTCCGGACG
151 GgcgAGATTG CGCCGTCCAA CCTGGATATCG GTCCGGCGCG AGGCCTTCGGG
201 GCAGTTAAA AAAGCTTATG TCAAACTGG GCAACAGTC AAAAAGGGCG
251 ATTGATGTC GGAATATCAAC TGCAACACGC AGACCAACAC GATCGATATG
301 GAAAATTCG AATTGGAAAC GTATCAGCGG AGACTGTGTG CGGCACAGAT
351 TGCAATTGGC AGCCGGGAGA AGAAATTAAG CGCTCAGGGC CGCTTGTGGA
401 AGGATGATGC GACCTCTAAAGAAGATTTGGG AAAGCGGCGA GGATGGGCTT
451 GCGCCGCGCA AACGCGATCT TGCGGAGTTG ARGGCTTAA TCAGACAGAG
501 CAAATTTCCG ATCAATACCG CGAGCTCGA TTTCGGCTAC ACGGCGATTA
551 CCCGAGCACAT GGACGGCAGC GTGGTGGCGA TCCCCGGGA AGAGGGGAG
601 ACTTGACCG CGCCGGCGTC TAGCGGACAG ATTTGCGGAT TGGCGAATCT
651 GGATATGATG TTGACACAAAG TGCAAGATTC CGAGGGGGAT ATTACCAAGG
701 TGAGGGCGC CGAGGATATT TCGATTTAGCA TTTTGTGCGA ACCGGATAG
751 CGGATTAAGC CGAACGTCGA CAGCGTCGAC CCCGGGCTGA CAACGATGTC
801 GTCGGGCGC CGAACGACGA GTACCGGATAC GGCTTCCAAAT GGCGTCTATT
851 ATTATGCCCG TTCTGTTGTC CGGAATCGGG AGCGCAACAT CGCCACGGGG
901 ATGAGCAGCG AGAAATACGGT TGAATATCGAC GTGTGAAAAA ATGTGTGCT
951 TATTCGGTC CTGACCGTC AAAATCGGG CGGCAAGGGC TTCTGACCG
1001 TGTTGGTGC GGACGGCGAG CGAGTGGAAC CGGAAATCCG GACCGGTATG
1051 AAAGACAGTC TGAATACCGA AGTGAAGAAC GGGTTGAAGA AGGGGGACAA
1101 ATG'TGTCATC TCGCAATTAAC CGCGCGCGGA CGAGCAGGAA AGCGCGAAC
25 1151 GCGCCCTAGG CGGGCCCGC CGCGATATA

```

This encodes a protein having amino acid sequence <SEQ ID 772>:

```

1 MAKKMKWVAAP AAVAAAIVWVG GWSYLKPEPQ AAYITEAVRR GDISRTVSAT
51 GEISPSNLVS VGRQAQSQIK KLYVQLQQV KKGDLLIAEIN STTQNTIDM
101 EKSKELETQY KLVSQAQIALG SAEKKYKRQKA ALWKDDATSK EDLESAQDAL
151 AAAAKANVAEL KALIRQSKIS INTAESDLGT TRITATMDGT VVAIPVEEGQ
201 TVNAQSTPP TIVQLNLDMM LNKMQTAEGD ITKVKAQODI SFTILSEPDFT
251 PIKAKLDSVD PGLTMMSSGG YNSTSTDASHE AVYYYYARSEV FPNPGKLATG
301 MTTQNTVEID GVKNVLILPS LTVKNRGGKA FVRVLGADGK AVEREIRTGM
351 KDSMNTEVKS GLKEGDKVVI SEITAEEQOE SCERALOGPP RR*

```

35 ORF85ng and ORF85-1 show 96.1% identity in 334 aa overlap:

	30	40	50	60	70	80	
orf85ng	PQAA	YITETV	RGRDIS	RTVSAT	GEI	SPSNLVS	GWSYLKPEPQ
							AAYITEAVRR
							GDISRTVSAT
orf85-1							
	90	100	110	120	130	140	
orf85ng	INSTTQNTIDME	KSKLETYQAKL	VLSA	QIALGS	AEKKYKRQ	ALWKD	ATSKEDLESQD
orf85-1	INSTSQTNTLNTE	KSLETYQAKL	VLSA	QIALGS	AEKKYKRQ	ALWKEN	ATSKEDLESQD
	40	50	60	70	80	90	
	150	160	170	180	190	200	
orf85ng	ALAAAKANVAEL	KALIRQSKIS	INTAES	DLGYTRIT	ATMDGT	VVAIPVEEGQ	TVNAAQST
orf85-1	ALAAAKANVAEL	KALIRQSKIS	INTAES	SELGYTRIT	ATMDGT	VAILVEEGQ	TVNAAQST
	100	110	120	130	140	150	
	210	220	230	240	250	260	
orf85ng	PTIVQLANLDM	MNLNKMQIA	EVDITKV	KAGQD	ISFTILSEPD	TPIKAKLDSVDP	GLTTMSS
orf85-1	PTIVQLANLDM	MNLNKMQIA	EVDITKV	KAGQD	ISFTILSEPD	TPIKAKLDSVDP	GLTTMSS
	160	170	180	190	200	210	
	270	280	290	300	310	320	
orf85ng	GGYN	SSTT	ASNAV	YYARS	FVNPDGK	LATGM	TQNTVEIDGVKNVLIPSLTVKNRG
orf85-1	GGYN	SSTT	ASNAV	YYARS	FVNPDGK	LATGM	TQNTVEIDGVKNVLIPSLTVKNRG
	220	230	240	250	260	270	
	330	340	350	360	370	380	

-431-

	orf85ng	KAFVRVRLGADGKAVEREIRTCMKDSMNTEVKSGLKEGDKVVISEITAEEQQESGERALGG : : : : : : : : : : : : orf85-1	280	290	300	310	320	330
5		390						
	orf85ng	PPRRX						
	orf85-1	PPRRX						

10 In addition, ORF85ng shows significant homology to an *E.coli* membrane fusion protein:

```
gi|1787104 (AE000189) c380; 27% identical (27 gaps) to 332 residues from
membrane fusion protein precursor, MTRC_NEIGO SW: P43505 (412 aa) [Escherichia
coli] Length = 380
Score = -193 bits (485), Expect = 2e-48
Identities = 120/345 (34%), Positives = 182/345 (51%), Gaps = 13/345 (3%)
Query: 29 PQRAYITEDLISRTVSATGEISPSNLIVSVAQGQASCQIKKKLYVKLCQQVKKGDLIAE 88
P T V R CD+ +V ATC++ V VQAQ SCQ+K L V +G +VKA L+
Sbjct: 41 PVTYTQTLIVRPGCDLQSQVLTGKLDALRKVWDGAQVSCQLKTLSVAIGDKVKKDQLLGV 100
20 Query: 89 INSTQTNTIMDEKSKLETYQKQLVSAQKALCSAEKKYKRQALWKDDTSKEXXXXXX 148
I+ N I ++ L +A+ A+ L A Y RQ L + A S+
Sbjct: 101 IDPEQAENQIKEVEATIMELRAQRQQAEEAELKLRVTVSRQQLAQTKAWSQOLDTAA 160
25 Query: 149 XXXXXXXXXXXXXXXXQSKISINTAESLGYTRITATMDGTVWAIPVEEGOTVNAQST 208
I++++ S++TA+++ L YTRI A M G V I +GOTV AAO
Sbjct: 161 EMAVKQAIQICTIDAQIKRNQASLDTAKTNLDYTRIVAPMAGEVTQITTLQGQTVIARQQA 220
30 Query: 209 PTIVOLANLDMLNKMQIAEGDIITKVKAGQDISFTILSEPDTPIKAKLDSVPDGLTTMS 268
P I+ LA++ ML K Q+E D+ +K GC FT+L +P T + ++ V P
Sbjct: 221 PNLTIDMSAMLVKAQVSEADWVHLPKGQKAWFTVLGDPFLTRYEGQIKDVLP----- 273
35 Query: 269 GCGYNSTDTASNAVYYARYSFVNPDPCKLATGMTTQNTVIEIDGVKNVLIPSLTVKNRGG 328
+ + ++A+YYAR VENF+L M MT Q +++ VRNVL IP + + G
Sbjct: 274 ----TEPKVNDAIYFYARFEVFNENGLRLRDMTAQWHQLTDVKNVLTIPLSALGDPVG 328
Query: 329 KAFVRV-LGADGKAVEREIRTCMKDSMNTEVKSGLKEGDKVVISE 372
+V L +G+ ERE+ G ++ + E+ GL+ GD+VVI E
Sbjct: 329 DNRYKVKKLRLNGETREPRETIGARNRNDTEIVKGLEAGDEVVIG
```

40 Based on this analysis, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF85-1 (40.4kDa) was cloned in the pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 19A shows the results of affinity purification of the GST-fusion protein. Purified GST-fusion protein 45 was used to immunise mice, whose sera were used for Western blot (Figure 19B), FACS analysis (Figure 19C), and ELISA (positive result). These experiments confirm that ORF85-1 is a surface-exposed protein, and that it is a useful immunogen.

Example 92

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 773>:

50	1	...ATTCGGCCCA CGATGACATT TCAACGAGC GGCAATGCTT ACAAAAATCGT
	51	TTCGAGGATT AAAGTCGGC TATACTATAT CCCTTTCGGAG TCCGGCGGTG
	101	CGGTCTCGG CAATACCTG CACCTTAACCT ACTATAGAGA CATACGCCAGC
	151	GCGCAACTCT ATCCGGAGnc CAAATTGGCC GAGcGAGC TAACCTACCG
	201	CAAAGCGGGC GAGACGCAAAG CCGGACAAAG CCCCAAGGCT ATGGATTCTG

-432-

251 TCAACGCTTGC CTGGCAGTTG GGGCGAAATG AGCGCGAAACT CCCCCCGGGG
 301 CTGAAAATCA CAAACGGCAA AAAACTTAT TCCTGTCGGG GTTGGATAA
 351 GGGGGGTACA GGAAAATACA GCATAGGGCG GAGCTGGAAACG GAAGTCGTCA
 401 AATATCGGGT GCGGGGGCG GACGATGCGG TAATGTTATTG TTTCGCAACCG
 451 TCCCTAACAA ATAATTCGGC ACAAAATGGC TATACCAGGC ACGCCAAAC
 501 CTATACGCTG AAACCTCAAAT CGGTGCGAGAT CARCGGCCAG GCAGCCAAAC
 551 CGTAA

This corresponds to the amino acid sequence <SEQ ID 774; ORF120>:

1 ..IPATMTFERS GNAYKIVSTI KVPLYNIRFE SGGTVVGNLH PTYYDRIRG
 51 GKLYAEAKFADGSVTYKGKAGESKTEQSPKAMDLFTLAWQL AANDAKLPGG
 101 LKTKNGKLKY SVGGLNKAGT KYSIGGVEET EVVKYRVRRRG DAVMYFFAP
 151 SLNNIPAQIG YTDDGKTYTLK LKSVQINGQ AKP*

Further work revealed the complete nucleotide sequence <SEQ ID 775>:

1 ATATGTAAGA CTTTTAAAAA TATATTTTC GCGGCCATIT TGTCGCCCGC
 51 CCTCCGGCTGC CGGTATCGG CAGGGCTGCC CCAATCCGGC GTGTCGCACT
 101 ATTCCGACAC CTACGGCATC CCGCCAGA TGACATTGGA ACCGAGGGC
 151 AATGCTCTACA AAATCGTTTC GACGATTAAG TGTCGGCTAT ACAATATCCG
 201 TTTCGAGTC GCGCTGAGCG TTTCGCGAC TACCCCTCAC CCTACCTACT
 251 ATAGAGACAT AGCGGAGGGC AAATCTGATAG CGGAAGGCCA ATTTCGGCCAC
 301 GGCAAGCTAA CTTCAGGCCAA AGCGGGCGG AGCAAAACCG AGCAAAAGCCC
 351 CAAGGCTTAGT GATTTGTTCA AGCTTGCGCTG GCAGTTGGCG CCAAATGAGC
 401 CGAAACTTCC CCCGGGGCTG AAAATCACCAC CGCGAAAAA ACTTTATTCC
 451 GTCCGGGGTT TGAATAGGG GGGTACAGGA AAATACAGCA TAGGGCGGGT
 501 GGCRAAGGAA GTGCTCAAT ATCCGGTGGC GCGCGCGAC GATCGCGTAA
 551 TGATTTCTT CGCACCTCC CTGAAACAATA TTCCGGACAA AATGGCTAT
 601 ACCGACGAGC GCAAAACCTA TAGCTGAACT CTCAARATCG TGCGAGATCAA
 651 CGGCCAGGCA GCGCAACCTA AA

This corresponds to the amino acid sequence <SEQ ID 776; ORF120-1>:

1 MMKTFKNIFS AATLSAALPC AVAAGLPQSA VLHYSGSYGI PATMTFERSG
 51 NAVKIVSTIK VPLVNIRFES GGTVVGNLH PTYYDRIRG KLYAEAKFAD
 101 GSYTGYKAGE SKTEQSPKAM DLFTLAWQL AANDAKLPGL KITNGKLYS
 151 VGGLNKAGTG KYSIGGVEET VVKYRVRRRG DAVMYFFAPS LNNIPAQIGY
 201 TDDGKTYTLK LKSVQINGQ AKP*

Computer analysis of this amino acid sequence gave the following results:

35 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF120 shows 92.4% identity over a 184aa overlap with an ORF (ORF120a) from strain A of *N. meningitidis*:

			10	20	30			
40	orf120.pep		IPATMTFERSGNAYKIVSTIKVPLYNIRFE					
	orf120a	SAATLSAALPCAYAAGLPLPXSAVLHYSGSYGI	:	:	:			
		10 20 30	40	50	60			
45	orf120.pep	SGGTVVGNLHPTYYDRIRRGKLYAEAKFADGSVTYKGKAGESKTEQSPKAMDLFTLAWQL	40	50	60	70	80	90
	orf120a	SGGTVVGNLHPTYYDRIRRGKLYAEAKFADGSVTYKGKAXXXXXXQSPKAMDLFTLAWQL	70	80	90	100	110	120
50	orf120.pep	AANDAKLPGLKITNGKLKYSGVGLNKGATGKYSIGGVEETVVKYRVRRGDDAVMYFFAP	100	110	120	130	140	150
	orf120a	AANDAKLPGLKITNGKLKYSGVGLNKGATGKYSIGGVEETVVKYRVRRGDDAVMYFFAP	130	140	150	160	170	180

55

-433-

		160	170	180
	orf120.pep	SLNNI PAQIGYTDGGKTYTLKLKSQVINGQAQKPX		
5	orf120a			
		190	200	210
				220

The complete length ORF120a nucleotide sequence <SEQ ID 77> is:

1	ATGATGAAGA	CTTTTAAAAAA	TATATTTTC	GCCGCCATT	TGTGCCGCC
5	CCTGCCGTCG	GGGTATCGC	CAGGGCTGC	CNAATCCGCC	GTGTCGCAC
10	ATTCCGGCAC	CTACCGCAT	CCCGCACNA	NNANNTNNGN	ACNNNGNC
15	AATTCGGCAC	AAATCGTTTC	GACGATTTAA	GTGCGCTAT	ACAATATCCG
20	TTTCGAGTC	GCGCTGCAAG	TTGTCGGCAA	TACCTGAC	CCTACCTACT
25	ATAGAGACAT	ACCGAGGGC	AAACTGTATG	CGGAAGGCCA	ATTGCGGCAC
30	GCGCAGGCTAA	CTTACGGCAA	AGCGNNNNNN	ACNNNNNNNC	NCGAACGCC
35	CAAGGCTATG	GTATTGTTCA	CGCTTCGCTNG	GCAGATTGGC	GCAAATGACG
40	CGAACATTCG	CCCGGGCGCT	AAATACTACCA	ACCGAAAAAA	ACTTATTTC
45	GTCGGGGTGT	TAAGTAAAGG	GGTAGACAGGA	AAATACAGCA	TAGCCGGGCT
50	GCGAACCGAA	GTGCTCAAAAT	ATCGGTGCG	GCGCGCGAC	GATGCGGTA
55	TGTATTCCTT	CGACGACCTT	CTGAGACATA	TTCCGCGACA	AATGCGCTAT
60	ACCGACGACG	GCAAAACCTA	TACGCTGAAA	CTCAAAATCGG	TGCAAGATCAA
65	GGCCAGGCGA	CCCCAACCGT	AA		

This encodes a protein having amino acid sequence <SEQ ID 77>:

1	MMMKTFKNIPS	AAILSAALPC	AYAAGLPDXA	VLHYSGSYGI	PATXXXXXXX
5	NAKKIVSTIK	VFLYNIRES	GSTTVGNTLH	PTTYRDIRRG	KLYAEAKFAD
10	GSVTYKGAXX	XXXQSXPAM	DLFILAWQLA	ANDAKLPPGL	KITNGKKLYS
15	VGGLNKAGTG	KYSIGVETE	VVKYRVRRGD	DAVMYFFAPS	LNNIPAQIGY
20	TDDGKTYTLK	LKSQVINGQA	AKP*		

ORF120a and ORF120-1 show 93.3% identity in 223 aa overlap:

	orf120a.pep	orf120-1	10	20	30	40	50	60
30	MMKTFKNIFSAAILSAALPCAYAAGLPSAVLHYSGSYGI	MMKTFKNIFSAAILSAALPCAYAAGLPSAVLHYSGSYGI	PATXXXXXXXXX	XNAXXXKIVSTIK				
			10	20	30	40	50	60
35	orf120a.pep	orf120-1	VPLYNIRFESEGGTVVGNTLH	PTTYRDIRRGKLY	AEAKFADGGSVTY	GKAXXXXXQSXPAM		
			70	80	90	100	110	120
40	orf120a.pep	orf120-1	VPLYNIRFESEGGTVVGNTLH	PTTYRDIRRGKLY	AEAKFADGGSVTY	GKAGESEKTEQSXPAM		
			70	80	90	100	110	120
45	orf120a.pep	orf120-1	DLFTLAWQLAANDAKLPPGLKITNGKLY	SVGGLNKAGT	GKYSIGGVETEVVKYVRGRD			
			130	140	150	160	170	180
50	orf120a.pep	orf120-1	DAVMYFFAPS	DAVMYFFAPS	DAVMYFFAPS	DAVMYFFAPS	DAVMYFFAPS	
			190	200	210	220		

Homology with a predicted ORF from *N.gonorrhoeae*

ORF120 shows 97.8% identity over 184 aa overlap with a predicted ORF (ORF120ng) from *N.gonorrhoeae*:

55	orf120.pep	IPATMTFERSGNAYKIVSTIKVPLYNIRFE	30
	orf120ng	SAAILSAALPCAYAARLPQSAVLHYSGSYGI	
60	orf120.pep	PATMTFERSGNAYKIVSTIKVPLYNIRFE	69
	orf120ng	SGGTVVGNTLHPPTYRDIRRGKLY	
	orf120ng	AEAKFADGGSVTY	
	orf120ng	GKAGESEKTEQSXPAM	90
	orf120ng	DLFTLAWQL	
	orf120ng	SGGTVVGNTLHPAYKD	
	orf120ng	IRRGKLY	
	orf120ng	AEAKFADGGSVTY	
	orf120ng	GKAGESEKTEQSXPAM	129
	orf120ng	DLFTLAWQL	

5	orf120.pep	AANDAKLPPGLKITNGKKLYS VGGLNKAGTGKYSIGGVETEVV KYVRVRGDDTVYFFAP	150
	orf120ng		189
	orf120.pep	SNNI PAQ IGY TDDGKTYT LKLKS VQING QAAKP	184
	orf120ng		223

The complete length ORF120ng nucleotide sequence <SEQ ID 779> is:

10	1 ATGATGAAAGA CTTTTAAAAA TATATTTTCC GCGGCCATT TTGCGCCGC
	51 CCTGGCGTGC GGGTATGCC CGAAGCTACCA CCAATCCGGC GTGCTGCATC
	101 ATTCCGGCAGT CTAAAGGCAATT CCCGGCCACGA TGACATTGTA AGCGAGCGGC
	151 ATAGCTTACA AAATGCTTTC GACGATTAAA GTGCCGTAT ACAATATCCG
	201 TTTCGAATCT CGCGCGTACGG TTTCGCGCAAC TACCCGTCAC CCTGCCTACT
	251 ATAAGACAT ACAGCAGGGC AAACCTGATAG CGGAAGGCAA ATTCCGCGAC
	301 GGCAAGCGTAA CCTACGGCA ACGGGGGCAG AGCAAAACCG AGCAAAAGCCC
	351 CAAGACTATG GATTGTTCA CGCTTGCGTC CGAGTTGGCG GCAATGACG
	401 CGAACATCCC CCGCGGGCTG AAAATCACCA ACGGCCAAAT ACTTTATCC
	451 GTGCGGGCCG TGAATAAGGC GGGTACGGGA AAATACAGCA TagggGGGT
	501 GGAACCGAA GTCTGCAAT ATCGGGTGGC GGGGGGGCAG GATAACGGTAA
	551 CGTATTCTTC CGCACCGCTC CTGAGAACATA ATACGGCTAT
	601 ACCGACGACO GCAAAACCTA TAGCTGAAG CTCAAATCGG TGCGAGATCAA
	651 CGGACAGGCC GCGCAACCGT AA

This encodes a protein having amino acid sequence <SEQ ID 780>:

25	1 MMTFKNIFSAAILSAALPC AYARLPOSA VLHYSGSYGI PATMTFERSG
	51 NAVKIVSTIK VELVNIRFES GGTVVGNLH PAYKDLIRG KLYAEAKFAD
	101 GSUTVGKAGE SKTEQSPKAM DLFTLAWQLA ANDAKLPPGL KITNGKKLYS
	151 VGGLNKAGTG KYSIGGVETE VVKYVRVRGD DTVTYFFAP S LNNI PAQIG*

30 In comparison with ORF120-1, ORF120ng shows 97.8% identity in 223 aa overlap:

35	orf120-1.pep 10 20 30 40 50 60
	MMTFKNIFSAAILSAALPCAYAAGLPSAVLHYSGSYGI PATMTFERSGNAYKIVSTIK
	orf120ng 10 20 30 40 50 60
	MMTFKNIFSAAILSAALPCAYAARLPSAVLHYSGSYGI PATMTFERSGNAYKIVSTIK
40	orf120-1.pep 70 80 90 100 110 120
	VPLYNIRFESESGTUVGNLHPIYYDRIRGKLYAEAKFADGSVTYKGKAGE SKTEQSPKAM
	orf120ng 70 80 90 100 110 120
	VPLYNIRFESESGTUVGNLHPIYYDRIRGKLYAEAKFADGSVTYKGKAGE SKTEQSPKAM
45	orf120-1.pep 130 140 150 160 170 180
	DLFTLAWQLA ANDAKLPPGLKITNGKKLYS VGGLNKAGTGKYSIGGVETEVVKYVRVRGD
	orf120ng 130 140 150 160 170 180
	DLFTLAWQLA ANDAKLPPGLKITNGKKLYS VGGLNKAGTGKYSIGGVETEVVKYVRVRGD
50	orf120-1.pep 190 200 210 220
	DAVMYFFAPSLNNI PAQIGY TDDGKTYT LKLKS VQING QAAKPX
	orf120ng 190 200 210 220
	DTVTYFFAPSLNNI PAQIGY TDDGKTYT LKLKS VQING QAAKPX

This analysis, including the presence of a putative leader sequence in the gonococcal protein

55 suggests that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 93

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 781>:

-435-

1 ATGTATCGGA GGAAAGGGCG GGGCATCAAG CGCTGGATGG GTGCCGGTGC
 51 .GGCTTTGCC GCCTTGGCT GCCTGGTTTG CGCGCTCGC GATACTTGA
 101 CTCCTTGTG GGTTGGCGG TGCTGGCGT ATGTATTGGR CCCTTTGGTC
 151 GAATGGTGC AGAAAAAAGG TTGTAACCGT GCATCCGCT CGATGCTGT
 201 GATGGTGTG TCCCTGATIT TGTTGTTGGC ATTATGTTG ATTATGTC
 251 CTAGCTGGT CGGGCAGTC AACATTTGG CATCCGGCT GCCCCAATTA
 301 ATCGGTTTA TGCGAACAC GCTGCTGCC TTGGTGAARA ATACATCGG
 351 CGGATATG TG GAATCGATC AGGCATCTAT TATTTGCTGG CTTCAGGGC
 401 ATACGGGAGA GTTGAACAC GGGCTTAAGG CGTGGTTTCC CGTGGTTGATG
 451 AGGCAGGGCG GCAATATT..

This corresponds to the amino acid sequence <SEQ ID 782; ORF121>:

1 MYRKGRGKPK PWMGAGXAKA ALWVILVFAIG DLTTPFAVAA VLAVYVLDPVLV
 51 EWLKQKGILNR ASASMSVNVF SLILLALLL IIVMLVQF NNLSRSLPQL
 101 IGFMQNTLLP WLKNNTIGGYV EIDQASIIAW LQAHGTGELSN ALKAWFPVLM
 151 RQGGNI..

Further work revealed the complete nucleotide sequence <SEQ ID 783>:

1 ATGTATCGGA GGAAAGGGCG GGGCATCAAG CGCTGGATGG GTGCCGGTGC
 51 GGCCTTTGCC GCCTTGGCT GCCTGGTTTG CGCGCTCGC GATACTTGA
 101 CTGGCTTGTG GGTTGGCGG TGCTGGCGT ATGTATTGGR CCCTTTGGTC
 151 GAATGGTGC AGAAAAAAGG TTGTAACCGT GCATCCGCT CGATGCTGT
 201 GATGGTGTG TCCCTGATIT TGTTGTTGGC ATTATGTTG ATTATGTC
 251 CTAGCTGGT CGGGCAGTC AACATTTGG CATCCGGCT GCCCCAATTA
 301 ATCGGTTTA TGCGAACAC GCTGCTGCC TTGGTGAARA ATACATCGG
 351 CGGATATG TG GAATCGATC AGGCATCTAT TATTTGCTGG CTTCAGGGC
 401 ATACGGGAGA GTTGAACAC GGGCTTAAGG CGTGGTTTCC CGTGGTTGAG
 451 AGGCAGGGCG GCAATATTG CAGCAGTATC GGCACACTTC GCTGCTTCC
 501 CTTGCTGTCT TACTATTTC TGCTGGATTG CGACGGGGTGC TGCTGGGGCA
 551 TTGGCCAAGT GGTCGGAGG CGTITTTGCC GTGCTTATAC GCGCATTACA
 601 GGCATTGTA ACAGGATTG GGGGAATT TTGCGCGGC AGCTTCTGGT
 651 AATCTGATG ATGGCTTGG TTAGCAGTTG GGGATTGGT CTGGTGGGGC
 701 TGGATCTGGG TTGTCCTAC GGTATCTTG CGGTATTTT GGTGTTTGT
 751 CCTTATCTGG GGCGCTTAC CGGATGTTG CTGGCCACCG TGCGUCCTT
 801 GCTCCAGTTC GGCTCTGGCA ACGGCTCATC ATCGGTTTGG GCGGTTTTG
 851 CGCTTGGAGA GTTTCCTCGA AGTTTTTCA TTACCGCGAA RTAGCTGGGA
 901 GACCGTTATG GGCTCTGCC TTGTTGGGT ATCTTTTGC TGATGGCGT
 951 CGGGCAGCTG ATGGCTTGGC CGGAATGTT GGGGGGATTC CTTTGGCG
 1001 CGGTAAACCTI GGCTCTGCTT CGCGAGGGG TGCAAAATAA TTTGCGGCC
 1051 AGTTTTIACC GGGCAGGTA G

This corresponds to the amino acid sequence <SEQ ID 784; ORF121-1>:

40 1 MYRKGRGKPK PWMGAGXAKA ALWVILVFAIG DLTTPFAVAA VLAVYVLDPVLV
 51 EWLKQKGILNR ASASMSVNVF SLILLALLL IIVMLVQF NNLSRSLPQL
 101 IGFMQNTLLP WLKNNTIGGYV EIDQASIIAW LQAHGTGELSN ALKAWFPVLM
 151 RQGGNI..SSNLLNLLI YFEFLDWWRQW SSGIAKILPVR RFAGAITTRIT
 201 GNLENVILGEF LRGQLLNL MGLVYGLGLV LVGLDSGPAT GMALGILFVEV
 251 PYLGAFTPLL LATVALLIQF GSWNGLSLSW AVFAVGQLE SFFITPKIVG
 301 DRIGLSPFWV IFSLMAFGQL MCFVGMLAGL PLAATLVLL REGVQKYFAG
 351 SFYGR*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

50 ORF121 shows 98.7% identity over a 156aa overlap with an ORF (ORF121a) from strain A of *N. meningitidis*:

orf121.pep	10	20	30	40	50	60
orf121a	MYRRKGGRGKIKPWMDAGAFAALVWLVLFGDTLT	PFVAAVLAYVLDPVLVEWLQKKGLNR				
	10	20	30	40	50	60
orf121.pep	70	80	90	100	110	120
orf121a	ASASMSVMVFSLILLALLIIVPMLVQFNNLSRSLPQLIGFMONTLLFWLKNNTIGGYV					

-436-

	orf121a	ASASMSVMVFSLILLALLIIVPMLVQFNNLASRLPQLIGFMQNTLLPWLNKNTIGGYV 70 80 90 100 110 120
5	orf121.pep	EIDQASIIIAWLQAHGTGELSNALKAWFPVLNRQGGNI 130 140 150
	orf121a	EIDQASIIIAWLQAHGTGELSNALKAWFPVLNRQGGNI VSSIGNLLLPLLLYFLDWRW 130 140 150 160 170 180
10	orf121a	SCGIAKLVFRRFAGAYTRITGNLNEVLGEFLRGQLLVMLIMGLVYGLGLVVLVGLDSGFAI 190 200 210 220 230 240

The complete length ORF121a nucleotide sequence <SEQ ID 785> is:

15	1	ATGTATCGCA GGAAGGGGC 6GGCATCAGG CCGTGGATGG ATGCCGGTGC 51 51 GCGGTGGCGC GCCTGGTTCT CGCGTCGGC GATACTTGAA 101 101 CTGGCTTGGC GGTTGGCGGC FGTCGGCGT ATGTATTGGA CCCTTGGTC 151 151 GAATGGTGC AGAAAAGGG TTTGACCGCT GCATCCGGTT CATGCTGT 201 201 GTATGGTGT TCTCTGATTT TGTTGTGGC ATTATGTGT ATTATGTCC 251 251 CTATGGTGT CGGGCAGTCA AACAATTTG CATCCGGCGCT GCCCAATTAA 301 301 ATCGGATTTA TGCAAGAACAC GCTCTGGCG TGTTTGAAGAA ATACAAATCGG 351 351 CGGATGATGTG GAATACGATC AGGCATCTTA TATTGGCTGG CTTCAGGGCG 401 401 ATAGCGGGCA TGAGGACAC GCGCTTAAGG CGTGGTTTCC CGTTTTGATG 451 451 AGGCAGGGCC GCAATGGTGT CAGCAATGATG GGCACCCCTGC TGCTGCTTCC 501 501 CTGGCTGCTT TACTATTCCC TGCTGGATTG GCACGGGTGG TGCTGGCGCA 551 551 TTGCAACATG GTTCCGAGG CGTTGGCGC GTGCTTATAC GGCGATTAA 601 601 GGCAATTGAG ACAGGAGTATG GGGCGAATTG TGCGCGGGC AGCTCTGGT 651 651 GATGCTGATTG ATGGGTTTGG TTACAGGGCTT GGGGGTTGGT CTGTCGGGC 701 701 TGGATTCGGG GTTGCATCTC GGTATGTTTT CGGTATTATT GTTTTTGTT 751 751 CCCTATTGGC CGCGCTTAC AGGACTGCTC CTGGCAACCG TCGCCGGCCT 801 801 GCTTCAGTTC GGTCTGGGA ACGGCATCTC GGTGTTGG CGGTTTTTG 851 851 CCGTAGGACA GTTCTGGAA AGTTTTTCA TTACGCCGAA AATGTTGGGA 901 901 GACGCTATGCG GCTCTGGCC GTTTTGCGGT ATCTTTGCG TGATGGCGT 951 951 CGGGCAGCTG ATGGGCTTTC CGGAAATGTT GGCGGGATTG CCTTTGGCG 1001 1001 CGGTCTGCTT CGGTCTGCTT CGCGAGGGC TGCGAAGATA TTTTGGCGC 1051 1051 AGTTTTACCGGGCAGGTA G
----	---	--

This encodes a protein having amino acid sequence <SEQ ID 786>:

40	1	MYRRKGRIK PWMMDAGAFAA ALVWLVFALG DTLTPFAVAA VLAYVLDPLV 51 51 EWLQKGGLNR ASASMSVMVFE LKLNKNTILL YFNLNTGGV NNLASRLPQL 101 101 IGFMQNTLLP WLKNTIGGYV EIDQASIIIAW LQAHGTGELSN ALKAWFPVL 151 151 RQGQVNSSI GNLLNPLLL YFNLWDWQRW SCGIAKLVPR RFAGAYTRIT 201 201 GNLINEVLFGEF LRQGLLWMLI MGIVLYGLIVV LVGLDSGFAI GMVAGILVVF 251 251 PYLGATFGLL LATVAALLQF GSWNGLLAWV AVFAVGQFLE SFFITPKIVG 301 301 DRIGLSPFWV FISLMNAFGQL MGFGVMLAGL PLAATLVLL REGVQKYFAG 351 351 SFYRGR*
----	---	---

45 ORF121a and ORF121-1 show 99.2% identity in 356 aa overlap:

	orf121a.pep	10 20 30 40 50 60 MYRRKGRIK PFWMDAGAFAAALWLVFALGDTLTPFAVAAVLAYVLDPLVEWLQKGGLNR
50	orf121-1	10 20 30 40 50 60 MYRRKGRIK PFWMDAGAFAAALWLVFALGDTLTPFAVAAVLAYVLDPLVEWLQKGGLNR
	orf121a.pep	70 80 90 100 110 120 ASASMSVMVFSLILLALLIIVPMLVQFNNLASRLPQLIGFMQNTLLPWLNKNTIGGYV
55	orf121-1	70 80 90 100 110 120 ASASMSVMVFSLILLALLIIVPMLVQFNNLASRLPQLIGFMQNTLLPWLNKNTIGGYV
	orf121a.pep	130 140 150 160 170 180 EIDQASIIIAWLQAHGTGELSNALKAWFPVLMBQGGNIVSSIGNLLLPLLLYFLDWRW
60	orf121-1	130 140 150 160 170 180 EIDQASIIIAWLQAHGTGELSNALKAWFPVLMRQGGNIVSSIGNLLLPLLLYFLDWRW
65	orf121a.pep	190 200 210 220 230 240 SCGIAKLVFRRFAGAYTRITGNLNEVLGEFLRGQLLVMLIMGLVYGLGLVVLVGLDSGFAI

	orf121-1	SCGIAKLVPPRFAGAYTRITGNLNLEVILGEFLRGQLLVMLIMGLVYGLGLVLVGLDSCFAI 190 200 210 220 230 240	
5	orf121a.pep	GMWAGILVFPYLGAFTGLLLATVAALLQFGSWNCILAVWAVFAVGQFLESFFITPKIVG 250 260 270 280 290 300	
	orf121-1	GMLAGILVFPYLGAFTGLLLATVAALLQFGSWNGILSVWAVFAVGQFLESFFITPKIVG 250 260 270 280 290 300	
10			
	orf121a.pep	DRIGLSPFWVIFSLNAMFGQLMGFVGMLAGLPLAAVTLVLLREGVQKYFAGSFYGRGX 310 320 330 340 350	
15	orf121-1	DRIGLSPFWVIFSLNAMFGQLMGFVGMLAGLPLAAVTLVLLREGVQKYFAGSFYGRGX 310 320 330 340 350	

Homology with a predicted ORF from *N.gonorrhoeae*

ORF121 shows 97.4% identity over a 156 aa overlap with a predicted ORF (ORF121ng) from *N.gonorrhoeae*:

20	orf121.pep	MYRRKGRRGIKPWWMAGAXAFAALVWLVFAIGDTLTTPFAVAAVILAYVLDPLVEWLQKGKLN 60
	orf121ng	MYRRKGRRGIKPWWMAGAXAFAALVWLVFAIGDTLTTPFAVAAVILAYVLDPLVEWLQKGKLN 60
25	orf121.pep	ASASMSVMVFSLLILLALLIIIPVMLVCGQFNNLASRLPOLIGFMQNTLLPWLKNTTGGV 120
	orf121ng	ASASMSVMVFSLLILLALLIIIPVMLVCGQFNNLASRLPOLIGFMQNTLLPWLKNTTGGV 120
	orf121.pep	EIDQASIIAWLQAHTEGELSNALKAWFVLMRQGGNI 156
30	orf121ng	EIDQASIIAWLQAHTEGELSNALKAWFVLMRQGGNI 180

An ORF121ng nucleotide sequence <SEQ ID 787> was predicted to encode a protein having amino acid sequence <SEQ ID 788>:

1	MYRRKGRRGIK PWWMAGAXAFA ALWVILVYALG DTLTTPFAVAA VLAYVLDPLV
51	EWLQKGKLN ASASMSVMVFS LLILLALLII IPVMLVCGQFNNLASRLPOLIGFMQNTLLPWLKNTTGGV
101	1IGFMQNTLLPWLKNTTGGV EIDQASIIAWLQAHTEGELSNALKAWFVLMRQGGNI
151	KQGNGNIVSTTL GNLLLPPLL YFPLLDWHRW SCGIPKLVPR RFAGAYTRIT
201	GNLNUKVGKF LRGQLLGTE RGAVVCRVGE ECWEGGGARS RPSDDGWPWR
251	GGG*

Further work revealed the following gonococcal DNA sequence <SEQ ID 789>:

40	1	ATGATATCGGA GAAAAGAACG GGGCATCAAG CGCTGGATGG GTGCCCGGGC
	51	GGCGTTTGC CCGCTTGGTC CGCTGGTTTA CGCCGCTCGGC GATACTTTGA
	101	CTCGCTTGC GTTTCGGCGC GTTCTGGCGT ATGCTGGTGA CCTCTTGCGC
	151	GAATGTTTGC AGAAAAGGGG TTGAAACCGT GCATCGCGCT CGATGCTGT
	201	GATGGTTTGC TCTCTGATT TTGTTGTTGG ATTATGTTG ATTATGTTG
	251	CTATGCTGT CGGGCGACTG AATAATTTGC CATCTGCCT GCCCCAAATA
	301	ATGCGTTTA TCAGAACACG TTGGCTGGCC TTGGTGAAAA ATACAATGG
	351	CGGATATGTC GAATCGATC AGGCATCTAT TATGGCGTGG TTTCAGGGC
	401	ATACGGGGCA GTTGACGACG GGCGCTTAAGG CTGCGTTTCG CGTTTTGATG
	451	AAACAGGGCG GCAATATGGT CAGCACTATC GCACACCTCG TGTGCGGCC
	501	CTTGTGCTTG TAATATTGTC TGCTGGATTC GCACCGGTG TGCTGCGCCA
	551	TGCCCAAACG GTTCCCGAGG CGTTTTGCCG GTGCTTATAC GCGCATTCAGC
	601	GGTAATTGCA ACAGGAGTAT GGCGGAATTTC TTGGCGCGTC AGCTCTCGGT
	651	GATGCTGATT ATGGCGCTTG TTACGGTTT GGGATTGATG CTAGTCGGAC
	701	TGGGATGGCG ATTTCGCAATG GGTAATGTTG CGGGTATTGTT GGTTGTTTGTG
	751	CCTTATTTGGG GTGGCTTGTAC GGGATCTGTC CTGGCACTG TTGCACTT
	801	GCTCCAGTTC GTTTCGGGAA AGCGGAATCTT GGCTGGTTGG CGCGTTTTTG
	851	CCCTCGCTCA GTTTCGGGAA AGCTTCTTCA TTACCCCGAA ATTTGATGGA
	901	GACCGTATCG GCCTGTCGCC GTTTGGGTT ATCTTTTCGC TGATGGCGTT
	951	CGGAAGAGCTG ATGGCTTGTG CGCGAATGTT GGCGGGATIG CCTTTGGCG
60	1001	CCGTAACACTT GTTCTGCTT CGCGAGGGCG CGCGAGAAAATA TTTTGGCCGC
	1051	AGTTTTTACG GGGCGAGGTA G

This corresponds to the amino acid sequence <SEQ ID 790; ORF121ng-1>:

	1	MYRRKGRIK PWMAGAGAFAA ALVWLVALYALG DTLTTPFAVAA VLAYVLDPVL
	51	EWLQKGGLNR ASASMSVMFE SLLILLALLL IIVPMVLVGQF NNLASRLPQL
5	101	IIGFMONTLLP WLKNNTIGGYV EIDQASIIAW FOAHGTGELSN ALKAWFPVLM
	151	KQGGNIVSSI GNLLLPFLLL YYFILDWORW SCGIAKLVP RFGAYTRIT
	201	GNLNEVLGEF LRGQLLVMLI MGLVYGLGIM LVGLDGGFAI GMVAGILVVF
	251	PYLGAFTGGL LATVAALLQF GSWNGLLAWW AVEFAVGQFLE SFFITPKIVG
	301	DRIGLSPEFWW IFSLMAFGEL MGFBGMLLAGL PLAAVTLVLL REGAQKYFAG
	351	SFYGR*

10 ORF121ng-1 and ORF121-1 show 97.5% identity in 356 aa overlap:

	10	20	30	40	50	60
	orf121-1.pep	MYRRKGRIK PWNMGAGAFAA ALVWLVALYALG DTLTTPFAVAA VLAYVLDPVL	WENLQKKGLNR			
15	orf121ng-1	MYRRKGRIK PWNMGAGAFAA WLVLVYALGDTLTPFAVAA VLAYVLDPVL	WENLQKKGLNR			
	10	20	30	40	50	60
	70	80	90	100	110	120
	orf121-1.pep	ASASMSVMVFSLLILLALLL IIVPMVLVGQFNNLASRLPQLIGFMONTLLPWLKNNTIGGYV				
20	orf121ng-1	ASASMSVMVFSLLILLALLL IIVPMVLVGQFNNLASRLPQLIGFMONTLLPWLKNNTIGGYV				
	70	80	90	100	110	120
	130	140	150	160	170	180
25	orf121-1.pep	EIDQASIIAWLQAHGTGELSNALKAWFVPLMVRQGGNIIVSSIGNLILLPLLYYFLLDWQRW				
	orf121ng-1	EIDQASIIAWFQAHGTGELSNALKAWFVPLMVKQGGNIIVSSIGNLILLPLLYYFLLDWQRW				
	130	140	150	160	170	180
30	orf121-1.pep	SCGIAKLVPFRFAFAGYTRITGNLNEVLGEFLRGQLLVNLMLGLVYGLGLVLGDLGSGFAI				
	orf121ng-1	SCGIAKLVPFRFAFAGYTRITGNLNEVLGEFLRGQLLVNLMLGLVYGLGLVLGDLGSGFAI				
	190	200	210	220	230	240
	250	260	270	280	290	300
35	orf121-1.pep	GMLAGILVFPYVPLGAGTGLLLATVAALLQFGSWNGILSVWAVEFAVGFQFLESFFITPKIVG				
	orf121ng-1	GMVAGILVFPYVPLGAGTGLLLATVAALLQFGSWNGILAVWAVFAVGQFLESFFITPKIVG				
	250	260	270	280	290	300
40	orf121-1.pep	DRIGLSPEFWVIFSLMAFGQLMGFVGMILAGLPLAAVTLVLREGVQKYFAGSFYGRX				
	orf121ng-1	DRIGLSPEFWVIFSLMAFGELMGFVGMILAGLPLAAVTLVLREGVQKYFAGSFYGRX				
45		310	320	330	340	350
		310	320	330	340	350

In addition, ORF121ng-1 shows homology to a permease from *H.influenzae*:

	sp P43969 PERM_HAEIN PUTATIVE PERMEASE PERM HOMOLOG Length = 349
	Score = 69.9 bits (168), Expect = 2e-11
	Identities = 67/317 (21%), Positives = 120/317 (37%), Gaps = 7/317 (2%)
50	Query: 26 VVALGDTLTPFAVAAVAVLVDPLPVLVEWL-QKKGLNRASASMSVMVFSXXXXXXXXXXVP 84
	+Y GD + P +R V+Y+L* + +L Q+ R A++ + VP
	Sbjct: 32 IYFGDLIAPLLIAVLSYLIEPIFLNQYIKCPMLTILILIGSFIGLAARVFFLVLVP 91
55	Query: 85 MLVGQFNNLASRLPQLIGFMQNTLLPWLKNNTIGGYVE-IDQASIIAWFQAHGTGELSNALK 143
	ML Q +L S LP + N WL N Y E ID + + F + ++ +
	Sbjct: 92 MLVGQFNNLASRLPQLIGFMQNTLLPWLKNNTIGGYVE-IDQASIIAWFQAHGTGELSNALK 143
	NL Q +L S LP + N WL N Y E ID + + F + ++ +
	Sbjct: 144 AWFPVLMKQGGNIIVSSIGNXXXXXXDXQWRWSGCAKLVPRPRAFAGYTRITGLN 203
60	+ + + N+VS D G+++ +P+ A+ R +
	Sbjct: 148 SAVKLSLASIMMLVSLIYAFLVLPMFFMLDKSELLQGSRFLFKNNRLAFXRWK-EM 206
65	Query: 204 NEVLFEGFLVPPGXXXXXXXXXXXXXXXXXXXXXXXDSGFAIGNVAGILVFPVYXXXXXXXXXXXX 263
	+ + ++ G+ + + + G+ V VPY
	Sbjct: 207 QQQISNYIHGKLLEILIVTLITYIIFLIGLNYPLLAFAVGLSVLVPYIGAVITIPVA 266

-439-

Query: 264 XXXXXQFGSWNGILAVWAVFNVQFLESFFITPKIVGDRIGLSPFWVIFSLMAFGELMGF 323
 QFG + FAV Q L+ + P + + + L P + I S++ FG L GF
 Sbjct: 267 LVALVFGFISPTFWYIIIAFVSQQLLDGNLLVPYLSEAVNLHPLIIIIISVLIFGGIWMGF 326

5 Query: 324 VGMIAGLPLAAVTLVLL 340
 G+ +FLA + ++
 Sbjct: 327 WGVEFAIPLATLVKAVI 343

Based on this analysis, including the presence of a putative leader sequence and transmembrane domains in the two proteins, it is predicted that the proteins from *N.meningitidis* and 10 *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 94

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 791>:

15 1 ..ACTGCTTTT CGGCAGCGCT GCGCTTGAGT CCATCATGAC TCGTCATATT
 51 51 TTTCGCTTCTT GGAAACCGCT ATCACACACG AGCCGCCATC TTACATTTT
 101 101 TTTCGACGTC CTGGCCCG CGTCAAACTG CTATCAGCA ATACCGCCG
 151 151 CTGGCCCTCT ATGGCTTCCA TCCGGCCGAG ATAGCCGAGT TTTCGTTGG
 201 201 TTTCGCTTT GATGTGGAC CAGCAAAATGT CTATGCCCA ATCGCGCCG
 251 251 ATGTTGGCAC GCATTTGCGG AATGTCGGG CGGAGTGTGG GTTTCCTGTC
 301 301 ATATCACGGTC GTATGACAT TGACGGCTC CCAACCTGC GCTGAACGC
 351 351 TTTCGATACCG CGCACCGAA AGGACGGCGG TGTCGGCATC TTGGAACCTCT
 401 401 CGCGGGGTGT CGGGGAAATG GCTGCGGATA TCGCCCAAAC CTGCGGACCC
 451 451 GAGCAGCGCC TCCTGTAACCG CGTGCAGCAG CGCATCGGCA TCGGAGTGTG
 501 501 CGAGCAGGCC TTTTCAAAT GGGATTCAA CTCCGCGAAC TATCAG..

25 This corresponds to the amino acid sequence <SEQ ID 792; ORF122>:

1 1 ..TAFSAALRLS FSXLVIFLSF GKPYQQTAII LTFFCTSCPP RSNAYQQYRR
 51 51 LRLYAFHPPF IAEEFVGFAF DVDAARNVYAQ IGGDVGTHLR NVRRECGLC
 101 101 NHGRIDIDR PTLRLNALIR RTQKDAAVRI FEIICGGVGM AADIAQTCRT
 151 151 EQQRVNGVQO RIGIGVSEQP FFKWDWFNSAK YO ..

30 Further work revealed the complete nucleotide sequence <SEQ ID 793>:

35 1 ATATCGTACT GGGCAAGCAG TTGCGCCGGAT TTTTTGAGG TAGATACCGC
 51 51 GCGCTTGTATI TTTTTGCGC CGTCTACCCAA GGCTCTGGATG AAAAAGTTGA
 101 101 TGGCTGAGGCC GGTCAGCTGG CGCGATATTATT CATTTCGGG TAGGATTTCG
 151 151 ACTGCTTTT CGGGCGGCGG CGCGCTTGAGT TCGCTCTGTG TCGTCATATT
 201 201 TTTCGCTTTT GGGAAACCGT ATCACACACG AGCCGCCATC TTACATTTT
 251 251 TTTCGAGCTG CTGGCCCGCG CGTTGCAARTG CTACCGACA ATACCGCCG
 301 301 CTGGCCCTCT ATGCTCTCA TCCGGCCGAG ATAGCGGAGT TTTCGTTGG
 351 351 TTTCGCTTTT GATGTGGAG CAGGAATGCT CTATGCCCA ATCGCGGGC
 401 401 ATGTTGGCAC GCATTTGCGG AATGTCGGG CGGAGTGTGG GTTTCCTGTC
 451 451 ATACACGGTC GTATGACAT TGACGGCTC CCAACCTGC GCTGAACGC
 501 501 TTTCGATACCG CGCACCGAA AGGACGGCGG TGTCGGCATC TTGGAACCTCT
 551 551 CGCGGGGTGT CGGGGAAATG GCTGCGGATA TCGCCCAAAC CTGCGGACCC
 601 601 GAGCAGCGCC TCCTGTAACCG CGTGCAGCAG CGCATCGGCA TCGGAGTGTG
 651 651 CGAGCAGCCC TTTTCAAAT GGGATTCAA CTCCGCGAAC TATCAGCTTT
 701 701 CTGCTTCTGG TCAGTTGGTG GACATCTAG CCTCTGTCCGA TACGGATGTT
 751 751 CGTCATCGTT TGTGTTCTGT A

This corresponds to the amino acid sequence <SEQ ID 794; ORF122-1>:

50 1 ISYWASSSSPD FLEVDTAPLII FLPILLPKASH KKLMLVEPVPM PIYSFSGTNS
 51 51 TAFSAAMRLS SSCWVIFLSF GKPYQQTAII LTFFCTSCPP RSNAYQQYRR
 101 101 LRLYAFHPPF IAEEFVGFAF DVDAARNVYAQ IGGDVGTHLR NVRRECGLC
 151 151 NHGRIDIDR PTLRLNALIR RTQKDAAVRI FEIICGGVGM AADIAQTCRT
 201 201 EQQRVNGVQO RIGIGVSEQP FFKWDWFNSAK YQLSAFGQLV DIVALSDTDV
 251 251 RHRRLCS*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N. meningitidis* (strain A)

ORF122 shows 94.0% identity over a 182aa overlap with an ORF (ORF122a) from strain A of *N. meningitidis*:

		10	20	30			
5	orf122.pep	TAFSAALRLSPSXLVIFL	SGKPYQQTAAI				
		:	:				
	orf122a	FPLLLPKASMKKLMVEPVPM	PMYSFSGTNSTAFSAAMRLS				
		30	40	50			
10	orf122.pep	40	50	60	70	80	90
		:	:	:	:	:	:
	orf122a	LTFXTSCPERSNPYQQYRRLYAFH	APEIIEFFVGFAF	DARNVYAQIGGDVGT	GTHLR		
		90	100	110	120	130	140
15	orf122.pep	100	110	120	130	140	150
		:	:	:	:	:	:
	orf122a	NVRRECGFLCNHGRIDLIDRLPTLRLNALIR	TQDAAVRIFELCGVGEM	AAIDIAGTCRT			
		150	160	170	180	190	200
20	orf122.pep	160	170	180			
		:	:	:			
	orf122a	EQRVGNGVQQRIGIVSEQPFKWDNFNSAKY	QLSASFQLVDIVALSDT	VRRHLCX			
		210	220	230	240	250	
25	orf122.pep	EQRVGNGVQQRIGIVSEQPFKWDNFNSAKY	QLSASFQLVDIVALSDT	VRRHLCX			

The complete length ORF122a nucleotide sequence <SEQ ID 795> is:

1	ATATCATATT	GGCAACGAG	TTCACTGGAT	TTTTTGAGG	TAGATACCGC
51	GCTTGTGATT	TTTTTGCCCC	TCTTACCCAA	GCCTTCCGATG	AAAAGCTTGA
101	TGGTCAACCC	GGTACCGATG	CGCATGTTATT	CTTTTTCCGG	TACGAATTCG
151	ACTGCNTTTT	CGGGCGGAT	GGCTTGTAGT	TGCTCTTGTG	TCTGCTATATT
201	TTTGTCTT	GGGAACCGT	ATCAACACAA	AGCCGGCCATC	TTAACATTTT
251	TTNNNACGTC	CTGGCCGCC	CGTTAACATC	CTTACCAAGCA	ATACCGCCG
301	CTGCGACTCT	ATGGCCCTCA	TGGGGCCGAC	ATAACCGAGT	TTTTCGTTGG
351	TTTTCGCTT	GANTTGAACG	CACGAATATG	CTATGCCAA	ATCGCGGGC
401	ATGTTGAC	CAATTGCGC	AATATGCGC	CGGAGTTGG	TTTCTCTGTC
451	AATCACGGTC	GTATGCACAT	TGACCGCCCTG	CCAACCTTCG	GCTCTGAACGC
501	TTTGATACCG	CCACGCAAA	AGGACCGCCGC	TGTCGGCATC	TTTGAACCTCT
551	CGCCGGTGT	CGGGAAATG	CGTGCCTGAT	TGCCCCAAAC	CTGCCGACC
601	GAGCACGCC	TCTGGTAACG	CGTGCAGCAG	CCCATCGGCA	TCCGGAGTGTG
651	CGAGCACGCC	TTTTTCAAT	GGGATTTCAA	CTCCGCCAAG	TATCAGCTT
701	CTGCGCTCGG	TCAGTTGGG	GACATCGTAG	CCCTGTCCGA	TACGGATGTT
751	CGTCATCGT	TCTGTTCTG	A		

This encodes a protein having amino acid sequence <SEQ ID 796>:

45	1	ISYWASSSSL	FLEVDTAPLI	FPLLPKASM	KKLMVEPVPM	PMYSFSGTN
51	TAFAA	MRLS	SSCVVIFLSF	GKPYQQTAAI	LTFXTSCP	RSPNPYQYQR
101	RLYAH	APEI	FFVGFAF	XVDA	RNVYAQ	IGGDVGT
151	NHGRIDLIDR	PTLRLNALIR	RTQDAAVR	FELCGVGEM	AAIDIAGTCRT	NMRRFGFLC
201	EQRVGNGVQ	RIGIGVSEQ	FFKWDNFNSA	YQLSASFQLV	DIVALS	STDV
251	RHLCS*					

ORF122a and ORF122-1 show 96.9% identity in 256 aa overlap:

		10	20	30	40	50	60
55	orf122a.pep	ISYWASSSSL	FLEVDTAPLI	FPLLPKASM	KKLMVEPVPM	PMYSFSGTN	STAFSAAMRLS
		:	:	:	:	:	:
	orf122-1	ISYWASSSPD	FLEVDTA	LFLPLL	PKASM	KKLMVEPVPM	PIYSFSGTNSTAFSAAMRLS
		10	20	30	40	50	60
60	orf122a.pep	70	80	90	100	110	120
		:	:	:	:	:	:
	orf122-1	SSCVVIFLS	GPKFYQQTAAI	LTFXTSCP	RSPNPYQYQR	RLYAH	APEI
		:	:	:	:	:	:
	orf122a.pep	SSCVVIFLS	GPKFYQQTAAI	LTFXTSCP	RSPNPYQYQR	RLYAH	APEI
		:	:	:	:	:	:
	orf122-1	SSCVVIFLS	GPKFYQQTAAI	LTFXTSCP	RSPNPYQYQR	RLYAH	APEI

-441-

The complete length ORF122ng nucleotide sequence <SEQ ID 797> is:

40	1	ATGTCG TACC	GGGA CAGC RG	TTC GCGG GT	TTT TGGAGG	TTGG AACCCG
	51	GCC TTTGATT	TTTTTACCGC	TTTT GCCCCA	GGCTTG ATGC	AAAGGATTA G
	101	tgTGC GAC	Gtta CGGATG	CGGA ATGT ATT	CTGTTT CGGG	TACCA AGTT CG
	151	ACTG CTTTT	CGGG GGGC AG	GGCG Ttc Agt	TGct gttt cgt	TGCT ATCAT ATT
	201	TTT Attc ct	gGGA Aaccct	atca aCaA Ac	ggccgc cat	TTAAC ATTTT
45	251	TTTG GAGC G	cttggccg ccg	cgttca AATC	cgttacca Gc	atccaggc Gc
	301	ctggccgt CT	Atgc CCTTCA	TTC GCGG CG	ATTC GCGG AT	TTT CGCT TGG
	351	TTTT GCTT	GATat TGAC G	CAGGA ATAT	CGat acCC Aa	atccgc gC G
	401	ATG TGGC AC	GCAT TTGCGG	AAT TGTGGG T	GGAG ATTT GG	GTTC TGTG C
	451	ATA CAGCG	GTAT CGCAC	TGACCC ATCT	CCACCC GT	GAAG GAC C
50	501	TTTG ATAGC G	CGCAGG CAGA	AGAG GGGC G	TTC GCGCAT	TTC GACTT
	551	GGCGGG GTG	CGGG AAAA ATG	GTG CGGAT G	TGG CCAA AT	CTGG CGC ACC
	601	GAGC CAGC cg	tggat tggaa	CGT GAGC AG	cggc tggt GCA	TTGG ATGCT
	651	CGCA GAGCC	TTTTT CAAT	GGG ATTT CCA	TCC CGGC AG	TATCC AGCTT
	701	CTGGCTT CGG	TCAT TGGT G	GAC ATCG TAG	CCCT GTCC GA	TACCGG ATTT
55	751	CGTC ATCT CG	TGT TGGT CTG	A		

This encodes a protein having amino acid sequence <SEQ ID 798>;

```

1 MSYRASSSPD FLEVETAPL FFLPLPMLK PPMKLMWPPM FPMFSSGTSN
5 TAFSAAMLBL GSCVVFISL PQQYQOATAI LTFCTCSWBLD RSNPYYQQR
10.1 LRLYAHFPEE TAEFFWGFBD DIDARHNDIQT IGGDVTGLH NVRECFGLC
15.1 NHGRIDIDHL PTLRLNALTQ RTQKDAAVRI EFGVTCGHL RGVNGVGNK
20.1 EQVRGNNGVQQ RVGTRMPQPF EFKWDFNSAK YQLSAFGQLV DVALSDTD
25.1 RHLRCLS

```

ORF122ng and ORF122-1 show 92.6% identity in 256 aa overlap:

		10	20	30	40	50	60
5	orf122-1.pep	ISYWASSSPDFLEVDTAPLLFLPLLPKASMKLMVEPVMPPIYSFSGTNSTAFSAMRLS					
	orf122ng	MSYRASSSSPDFFLEVETAPLLFLPLLPKASMKLMVEPVMPMYSFSGTNSTAFSAMRLS	10	20	30	40	50
10	orf122-1.pep		70	80	90	100	110
	orf122ng	SSCVVIFLSFGKPYQQTAAILTFCTSCPERSNAYQQYRRLRLYAFHPEIAREFFVGFAF	70	80	90	100	110
15	orf122-1.pep		130	140	150	160	170
	orf122ng	DVDARNVVAQIGGDVGTHLNRVRREFGFLCNHGRIDIDLRLPTLRLNALIRRTQKDAAVRI	130	140	150	160	180
20	orf122-1.pep		190	200	210	220	230
	orf122ng	FELCGGGVGEAADIAQTCRTEQRVGNQVQORIGIVSEQFFFWDNFNSAKYQLSAFGQLV	190	200	210	220	240
25	orf122-1.pep		250				
	orf122ng	DIVALSDDTDVRHLCSX	250				
30							

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 95

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 799>:

35	1	..GCCGGCGGA	GTGGCAACAA	CATTTCCCGG	CGTTTGCGG	AAACACCCGT
	51	CGCTGTCAGC	CTTACCCCTGA	TGGCCACGGG	ACTTGCCTGC	ATGCTGGCGG
	101	TTACCGATAA	TGAANACCTC	TCGCTCTTA	TGGCTCGGG	ATTGGCCGG
	151	ATGGCGCGGA	TTTTGATTGC	CGACTTTTC	GCTTGTAAAC	GGCGTGA

This corresponds to the amino acid sequence <SEQ ID 800; ORF125>:

40	1	.AGASANNISA	RFAETPVAVS	VTLIGTVLAV	MLPVTEYENF	LLLIGSVFAP
	51	MGGFDCLFR	LETA*			

Further work revealed the complete nucleotide sequence <SEQ ID 801>:

45	1	ATGTCGGGCA	ATGCCCTCTC	TCCCTCATCT	TCCCTCCGCCA	TGGGGCTGAT
	51	TTGGTTCCGC	GGGGCGCTAT	CGATTGCGGA	AATCAGCACCG	GGTAGCCGCTGC
	101	TTGCCGCCTTC	GGGGCTGCCAG	CGGGGTCTGG	CGGCTTCTACT	TTGGGTCTAT
	151	GGCGTCGGCC	GGGGCGCTGT	TTTGGCGGGC	CGCTATATTCG	GGCCACTGAC
	201	CGGAGCCAGC	TCGATGGAAA	GTGTCGCCG	GTGTTGCGGC	AAACAGGGTT
	251	CAGTGGCTGTT	TTCCGTTGGCC	AATATGCTGCC	AACTGGCCGG	CTGGACGGCG
	301	GTGATGATTTC	ACGGCGCCGCC	ACCGTCAGC	TCCGCTTGG	GCAGAAGTGT
	351	GTGGACCGCC	GRATCTTTTG	TCGGTGGGGC	ATTCGGCAAC	GGGGCCGCTGA
	401	TTGTGCTGTC	GCTGTTTTTC	GGGGCACGCCA	AAACAGGGCG	CTGTAAGAAC
	451	GTTCGATGTC	GTTCGGGGTT	CTGTCGCTGA	GTGCGGAAGT	
	501	CTTTCACCG	GCAGCCGCCG	CGCCGCCACCA	GTTCAGACAG	GGCATGAGTT
	551	TCCGAAACGCC	ACTCGACGTC	TCCGGCGTGA	TGGCGCTTTC	CTGGCTGGCG
	601	CTTGGCCGCCG	ACTACACGCC	CCACGCCGCC	CGCCCGTTTG	CGGCACCCCT
	651	GACGGCAACG	CTGCCTACAC	CGTGTACCGG	CTGCTGGATG	TATGCCCTGG
	701	GTTCGGCGAC	GGCGTGTGTC	ACCGAGGAAA	CGGACGTGCC	AAAAATCCCTG
	751	CTGGGCGCAG	GTTCGGGTGTC	GGCAGGGCAT	TTGGCGGTGCG	TCTCTCCAC

801 CGTTTACCAAC ACGTTTCCTG ATGCCCTATTG CGCCGGCGCG AGTGGAAACA
 851 ACATTCGGC CGGTGTTGGG GAAACACCGC TCGCTGCG CGTGTACCTG
 901 ATCGGCACGG TACTTGCCTG CATGCTGCCG GTTACCGAAT ATGAAAACCTT
 951 CCTGCTGCTT ATTCGCTCGG TATTGGGCCGATGGCCGG GTTTGATTG
 5 1001 CGGACTTTT CGTCTTGAAGA CGGCGTGAAG AGATTGAAGG CTTTGACTTT
 1051 CGCGGACTGG TTCTGGGGT TGCGGGCTTC ATCTCTTACCG CTCTCTCTG
 1101 CTGTCGGC GGGAAAGCA GCATGGCT GACCGCCCCC GTAAATGCTG
 1151 CGCTTGCCTAT GGCAACCGTA TCGGTACGCC TTTCTTTAA AAAACCCCAA
 1201 TCTTACAAA GGAACCCGTC ATGA

- 10 This corresponds to the amino acid sequence <SEQ ID 802; ORF125-1>:

1 MSGNASSPSS SSAIGLINVFG AAVSIAEIST GTLLAPLGWQ RGLAALLLGH
 51 AVGGALFFAA AVIGALTGRS SMESVRLPSG KRGSVLFVSA NMQLLAGWTA
 101 VMIYAGATVS SALGVKLWMD ESEFWVWALAN GALIVLMLVFL GARAKTGGLKT
 151 VSMMLLMLAV LWLWSAEVPSL AGSTAAQVSD GMSFGTAVEL SAVMPJLSWLP
 201 LAADYTRHAR RPPFAATITLAY LAYLTGCMW YALGLAAALF TGETDVKIL
 251 LGAGLGMAGI LAVLVLSTVT TFLDAYSAGA SANNISARFA ETTFVAVGVT
 301 IGTVLAVMLVE ITBYEENFLLIGSVFAPMAA VLIADFFVLUK RREEIEGDFD
 351 AGLVLMVLAGE ILYRFLLESS WESSIGLTAB VMSAVAIATV SVRLFFKKTQ
 401 SLQRNPNS*

- 20 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N. meningitidis* (strain A)

ORF125 shows 76.5% identity over a 51aa overlap with an ORF (ORF125a) from strain A of *N. meningitidis*:

25	orf125.pep	10 20 30
		AGASANNISARFAETPVAVSVTLIGTVLAV
		::: ::: ::: ::: :::
	orf125a	KILLGAGLGAAGILAVVLSTVTTFLDAYSAGVSANNISAKLSEIPIAVAVAVVGTLLAV
		250 260 270 280 290 300
30	orf125.pep	40 50 60
		MLPVTEYENFLLLIGSVFAPMGGFDCLRFLETAX
		::: ::: ::: ::: :::
	orf125a	LLPVTEYENFLLLIGSVFAPMMAAVLIADEFFVLUKRREEIEG
		310 320 330 340

- 35 The ORF125a partial nucleotide sequence <SEQ ID 803> is:

1 ATGTCGGGCA ATGCCCTCTC TCNTTTCATCT TCCGCCGCCA TCGGGCTGAT
 51 TTGGTTCGGC CGGGCGGTAT CGATTGCGGA ATTCAGCAGC GGTCACACTGC
 101 TTGCGCCCTT GGCGCTGCCAG CGCGGTCTGG CNCGCTCTGT TTGGGGTCTAT
 151 GCGGTGCGGC GGGGGCTGTT TTGGGGGGCG CGCTGATATCGC GGCACATGAC
 201 CGGACANCANC TCGATGGAA CGGTGCGCCCT GTGCTTGGCG AACACCGGTT
 251 CAGTGCTGTTT TTGGCGGC AATCTGTCG AACTGCGCCCG CTGGACCGGG
 301 GTTGATGATTG ACCGGCGCGC AACAGCTCAGG TCGCGCTTGG GCAAAGTGT
 351 GTGGGACGTTT CGCTGGTTTG GCGCAGCGCA AAACAGCGCG GCTGAAAAACC
 401 GTTGTCGCTG TGCTGGTTTG GCGCAGCGCA AAACAGCGCG GCTGAAAAACC
 451 GTTGGCATGGC TCGTGTGGGT GTTGGCGGTG TTGTCGCTGAA GTGCCGAANT
 501 NTTTCCACG CGACGAGCGA CGCGGCGCAGG GTTNNCAGAC GGCATGAGTT
 551 TCGGAAACGGC AGTCGAGCTCA TCCGGCGTNA TGCGGCTTTC TTGGCTGCCG
 601 CTGGCCGCC ACTACACGG CCACCGCGG CGGCCCTTGG CGGCCAACCC
 651 GACGCCAACG CTGGCCCTAAC CGCTGACCGG CTGCTGGATG TTGCGCTTGG
 701 GTTGGCAGC CGCGCTGGTC ACCGGAGAAA CGCACCTGGC AAAATCCTG
 751 CTGGGGCGAG GTTGGGGTGC GGCAAGGCTT TTGGCGGTG TGCTGTCGAC
 801 CGTTRACACC ATCTTCTCTG ATGCTACTG CGCCGGCTGA ATGCGCAACA
 851 ATATTTCACG CAAACATTCG GAATATACNA TCGCCGTG CGTCGCCGTT
 901 GCTGGCACAC TCGTGTGGCT CCTGGCTGCCG GTTACCGGAT ATGAAAACCTT
 951 CGCTGCTGTT ATCGCTGGCG TATTGGCGG GATGGCGGCC GTTGGTATTG
 1001 CGGACTTTT CGTCTTGAAG CGGCGTGAAG AGATTGAAGG C..

This encodes a protein having the partial amino acid sequence <SEQ ID 804>:

1 MSGNASSXSS SAAIGLINVFG AAVSIAEIST GTLLAPLGWQ RGLAALLLGH
 51 AVGGALFFAA AYIGALTGXG SMESVRLPSG KRGSVLFVSA NMQLLAGWTA

101 VMIYAGATVS SALGKVLWLG ESEFWWWALAN GALIVLWLVE GARKTGGLKT
 151 VSMLLMLLAV LWLSAEXEST AGSTAAXVXD GMSEFTAVEL SAVMPLSLWP
 201 LAADYTHAR RPEFAATLTAT LAYLTGCGW YALGLAAALF TGETDVKAL
 251 LGAGLGAGGI LALVLSVTI TFTLDAYSAGV SANNISAKLS EPIPIAVAVAV
 301 VGTLLAVLPP VTEYENFILL IGSVFAMMA VLIAADFVLLK RREEIEG..

ORF125a and ORF125-1 show 94.5% identity in 347 aa overlap:

	10	20	30	40	50	60
10	orf125a.pep	MSGNASSSSAIIGLIWFQGAASVIAEISTGTLLAPLGWQRGLAALLLGHAVGGALFFAA	: : : : : : : : : : : :			
	orf125-1	MMSGNASSPSSAIIGLIWFQGAASVIAEISTGTLLAPLGWQRGLAALLLGHAVGGALFFAA	: : : : : : : : : : : :			
		10 20 30 40 50 60				
15	orf125a.pep	AYGALGTGXSMESVRLSFGKRGSVLFVANMLQLAGWTAVMIYAGATVSSALGKVWLWDG	: : : : : : : : : : : :			
	orf125-1	AYIGALTGRRSSMESVRLSFGKRGSVLFVANMLQLAGWTAVMIYAGATVSSALGKVWLWDG	: : : : : : : : : : : :			
		70 80 90 100 110 120				
20	orf125a.pep	ESFVWWALANGALIVLWLVLFGARKTGGLKTVSMLLMLLAVLWLSAEFSTAGSTAAXVD	: : : : : : : : : : : :			
	orf125-1	ESFVWWALANGALIVLWLVLFGARKTGGLKTVSMLLMLLAVLWLSAEFSTAGSTAAXVD	: : : : : : : : : : : :			
		130 140 150 160 170 180				
25	orf125a.pep	GMSFGTAVELSAVMPLSWSLPLAADDYTRHARRPFAATLTAFPLAYTLTGCWNMYALGLAAALF	: : : : : : : : : : : :			
	orf125-1	GMSFGTAVELSAVMPLSWSLPLAADDYTRHARRPFAATLTAFPLAYTLTGCWNMYALGLAAALF	: : : : : : : : : : : :			
		190 200 210 220 230 240				
30	orf125a.pep	TGSTDVAKILLGAGLGAAGILAVLVLSTVTTTFLDAYSAGVSANNISAKLSEIPIAVAVAV	: : : : : : : : : : : :			
	orf125-1	TGSTDVAKILLGAGLGAAGILAVLVLSTVTTTFLDAYSAGASANNISARFAETPVAVGVTL	: : : : : : : : : : : :			
		250 260 270 280 290 300				
35	orf125a.pep	VGTLLAVLPLPVTEYENFILLIGSVFAPMAVLLIADFVLLKRREEIEG	: : : : : : : : : : : :			
	orf125-1	IGTVLAVLPLPVTEYENFILLIGSVFAPMAVLLIADFVLLKRREEIEGPDFAGLVLWLAF	: : : : : : : : : : : :			
		310 320 330 340 350 360				

Homology with a predicted ORF from *N.gonorrhoeae*

45 ORF125 shows 86.2% identity over a 65aa overlap with a predicted ORF (ORF125ng) from *N.gonorrhoeae*:

	orf125.pep	AGASANNISARFAETPVAVSVTLLGTVLVA	30
		: : : : : :	
50	orf125ng	KILLGAGLGITGILAVVLSTVTTTFLDTSAGASANNISARFAEIPVAVGTVLIRTVLVA	308
		: : : : : : : :	
	orf125.pep	MLPVTEYENFILLIGSVFAPM-GGFDCRLFRLETA	64
		: : : : : : : :	
	orf125ng	MLPVTEYKNFILLIRSVEFGPMAGGFDCRLFCLETA	343
		: : : : : : : :	

An ORF125ng nucleotide sequence <SEQ ID 805> was predicted to encode a protein having amino acid sequence <SEQ ID 806>:

	1	M3GNASSPSS AAIIGLWFG AAVSIAEIST GTLLAPLGWQ RGLAALLLGH	
	51	AVGGALEFRA AYIGALTRGS SNESVRLSPG KCGSVLFSVVA NNOLQAGWTA	
101	VMIVYAGATVS SALGKVLWLG ESEFWWWALAN GALIVLWLVE GARRTGGLKT		
151	VSMLLMLLAV LWLSAEXEST AGSTAAXVXD GMSEFTAVEL SAVMPLSLWP		
201	LAADYTHAR RPEFAATLTAT LAYLTGCGW YALGLAAALF TGETDVKAL		
251	PLAADDYTRQA RRFPAATLTAT LAYLTGCGW MYALGLAAALF FTGETDVKAL		
301	LLGAGLGITG ILALVLSVTI TFTLDTYSAG ASANNISARF AIIPAVVGVT		
		LIRTVLAVLPL PVTEYKNFILL IRSVEFGPMAGGFDCRLFCLETA*	

Further work revealed the following gonococcal DNA sequence <SEQ ID 807>:

```

      1 ATGTCGGGCA ATGCCCTCTC TCCCTCATCT TCCGCCGCCA TCGGGCTGGT
      5 TTGGTCGGC GCGCGGTAT CGATTGCCGA AATCAGCAAGS GTTACGCTGC
10 101 TCGCCCTCTT CGGCTGCCAG CGCGCTCTGG CGGGCTCTGT TTGGGTCTAT
15 151 GCGCTGGCG CGCGCTCTGT TTGGGGCGG CGGTATATCG CGCACTGAC
20 201 CGGACGCCAGC TGATGGAAA GTGTCGGCT GTCTGGCC AAATGGGGT
25 251 CAGTGTGTT TTCCGTGGCG AATATGCTGC AACTGGCCGG CTGGACGGGG
30 301 GTGATGATT AGCTCGGCCG AACGGTCAGC TCCGCTTGGG AAAGGTGTT
35 351 GTGGGACGGC GAATCTTGT TCTGGTGGGC ATGGCAAAC GGCGCACTGA
40 401 TCGTGGCTGTG GCTGTTTGC GGCGCACGCA AAACGGGGCG CCTGAAAACC
45 451 GTTTCGATGC TCGTGTGCTG GCTGCTGGT TTGTGGTGA CGTGTGAAAGT
50 501 GTTCTGGTGC TCCGTCATCAA ACGCCCGCGC CGCCGTTTCA GACGGCATGA
55 551 CTCCTGGAGC GCGAGCTGAA CTGTCGGCC TCACTGGCGT TTCTGGCTG
60 601 CCCGCTGGCC CGCGATCACG CGCCCAAGCA CGCCGCCCGT TTGGGGCAAC
65 651 CTCGCTGGCA ACGACGCSCT ATACGCTGA GGGCTGCTGG ATGATGCTT
70 701 TGGGGTGGC CGCGCTCTGT TTACCGGG AAAACGGACGT GGCGAAAATTC
75 751 CTGTTGGCC CGGGCTCTGG CATAACGGG ATTCTGCGAG TCGTCTCTC
80 801 CACCGTACAC ACAAATGTTT TCAGTACCTA TTCCGCGGGC GGCAAGTGGCA
85 851 AACACATTTC CGCGCTGGT CGGAAATAC CGCTGCTGT CGCGCTTAC
90 901 CTGATGGCC CGGTGCTGC CGCTATGCTG CGCGTGTACCG AAATATAAAA
95 951 CTTCCTGCTG CTTATCGCTG CGGTATTTGC GCGCATGGG CGGGTTTGG
100 1001 TTGGCGACTT TTTCGCTTA RRPFAAFLTA TPLAYLTGCG MYALGLAAL FTGETOVAKI
105 1051 TTGGCGGAC TGGTCTGGT GCTGCGAGC TTCACTCTC ACCGCTTCTC
110 1101 GCTCTGSTC GGTGGGAAA GCAGCATCGG TCTGACCGCC CCGTAATGT
115 1151 CTGCGTGTG CATTGCCACC GTATCGTAC GCCTTTCTT TAAAAAAACC
120 1201 CAATTCTTAC AAAGAACCA CGTATGTA

```

This corresponds to the amino acid sequence <SEQ ID 808; ORF125ng-1>:

```

      1 MSGNASSPSS SAAIGLWFG AAVSIAEIST GTLLAPLGWQ RGLAALLLGH
      5 AVGQLFFAAV AVIGALTFGRS SMESVRLSLFG KCGSVLFSVA NMHQLAGWTA
10 101 VMIYVGATVS SALGKVLWDG ESFVWALAN GALIVLMLVF GARTRGKLK
15 151 VSMLLMLLAV LWLSEVFAS SGTNAAAPAVS DMGTFGTAVAE LSAMPLSLW
20 201 PLADYTRQAA RRPFQAFLTA TPLAYLTGCG MYALGLAAL FTGETOVAKI
25 251 LLGAGLGTIG LLAVLVLTVT TPLDLYTSAQ ASANNISARF AEIPVAVGVT
30 301 LIGTVLAVML PVTEYKNEFL LIGSVFAPMA AVLIADFVYL KRKEEIEGFD
35 351 FASLWLMLAG FILYRFLLSS GWESSIGLTA PVMASVAIAT VSVRLFFKKT
40 401 QSLQRNP5*

```

ORF125ng-1 and ORF125-1 show 95.1% identity in 408 aa overlap:

	10	20	30	40	50	60
40 orf125-1.pep	MSGNASSPSSSSAAGLIGLWF	GAASVIAEISTGTLLAPLGWQ	RGLAALLLGHAVGGALFFAA			
orf125ng-1	: : : :	: : : :	: : : :			
	10	20	30	40	50	60
45 orf125-1.pep	AYIGALTGRSSMESVRLSF	GKRGKSVLSVANMLOLAGWTAV	MIYAGATVSSALGKVLWDG			
orf125ng-1	: : : :	: : : :	: : : :			
	70	80	90	100	110	120
50 orf125-1.pep	ESEFWWALANGALIVLWLV	FGARKTGKTVSMLLMLLAVLWLSAEVFSTAGSTAQQ-VS				
orf125ng-1	: : : :	: : : :	: : : :			
	130	140	150	160	170	179
55 orf125-1.pep	DGMSPFGTAVELSAVMPLS	WLPLAADYTRHARRPFAATLT	ATLAYLTGCGMYALGLAAL			
orf125ng-1	: : : :	: : : :	: : : :			
	140	160	160	170	180	
60 orf125-1.pep	180 DGMFTGTA	VELSAVMPLSWLPLAADYTRHARRPFAATLT	ATLAYLTGCGMYALGLAAL			
orf125ng-1	: : : :	: : : :	: : : :			
	190	200	210	220	230	239
65 orf125-1.pep	FTGETDVAKILLGAGL	GAGLTLAVVLTSTVTFLDAYSAGASANNISARFAETPVAVGVT				
orf125ng-1	: : : :	: : : :	: : : :			
	250	260	270	280	290	299

		300	310	320	330	340	350	359
	orf125-1.pep	LIGTVLAVMLPVTEYENFLLIGSVFAPMAAVLIRADFVLRKREEIEGDFAGLVLWLAG						
5	orf125ng-1	LIGTVLAVMLPVTEYKNFLLIGSVFAPMAAVLIRADFVLRKREEIEGDFAGLVLWLAG						
		310	320	330	340	350	359	360
	orf125-1.pep	360	370	380	390	400		
10	orf125ng-1	FILYRFLLSSGEWESSIGLTAPVMSAVAIATSVRLFFKKTQSLQRNPSX						
		370	380	390	400			

Based on this analysis, including the presence of putative leader sequence and transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 96

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 809>:

20	1	ATGACCCGTA	TGCCATCCT	CGCGGGCGGC	CTCTCGGGA	GGCTGACCCG		
	51	GTTCGACGTT	CGAACACAA	GTTATCAGAT	TGCACTTTTC	GAATAAAGCT		
	101	GGCCGGGGGG	CGAACACCGC	GGCGCTATG	TAGGCCGCC	CATGCTCCG		
	151	CTCTGAGCGGS	AAGCTCGCA	AGCCACGCC	GAAGTGGTCA	GGTGGGAG		
	201	GCAGAGCATC	CGCTTITGG	GGCGCATCC	ATGCCGCTG	AACACGGCAC		
	251	CGATGATC	CGAAAACGG	AGCTGATTC	TATGCCACG	GCAGGACAA		
	301	CCATTATC	CGGAGTTCT	CGCCGATCTC	AAACGGCGG	GCCTTACCGA		
	351	TGACGAAATC	GTCCGTTGG	GGCCGACGA	CATGCCGAA	CGCGAACCG		
	401	AACTCGGG	ACGTTTTG	GGCGGCATCT	ACCTGGCGAC	CGAAGC	CAG	
	451	CTCGACGGG	GGCAATTATA	GTCTGACTT	GGCGACGCTT	TGGACGAACT		
	501	GAACGCTCCC	TGCCATTG	GACAGGAT	CSTCCCGAA	GGCGTCGAG		

30 This corresponds to the amino acid sequence <SEQ ID 810; ORF126>:

1	MTRAILGGG	LSGRLLTALQL	AEGQYQIALF	DKSCRRLGEHA	AAYVAAAMLA		
51	PAAXTVEATP	EVVRLGRQSI	PLWRGIRCRRL	NHTHMMQENG	SLIVVWHGDK		
101	PLSSEFVRHL	KRGKXTDIE	VRWRADDIAE	REPOLGGRFX	DGIYLPTEXQ		
151	LDGRQLKSAL	ADALDELNV	CHWEHECVPE	ACK...			

35 Further work revealed the complete nucleotide sequence <SEQ ID 811>:

1	ATGACCCGTA	TGCCATCCT	CGCGGGCGGC	CTCTCGGGA	GGCTGACCCG		
51	GTTCGACGTT	CGAACACAA	GTTATCAGAT	TGCACTTTTC	GAATAAAGCT		
101	GGCCGGGGGG	CGAACACCGC	GGCGCTATG	TTGCCGCC	CATGCTCCG		
151	CTCTGAGCGGS	AAGCTCGCA	AGCCACGCC	GAAGTGGTCA	GGTGGGAG		
201	GCAGAGCATC	CGCTTITGG	GGCGCATCC	ATGCCGCTG	AACACGGACA		
251	CGATGATC	CGAAAACGG	AGCTGATTC	TATGCCACG	GCAGGACAA		
301	CCATTATC	CGGAGTTCT	CGCCGATCTC	AAACGGCGG	GCCTTACCGA		
351	TGACGAAATC	GTCCGTTGG	GGCCGACGA	CATGCCGAA	CGCGAACCGC		
401	AACTCGGG	ACGTTTTG	GACGGCATCT	ACCTGGCGAC	CGAAGGGAG		
451	CTCGACGGG	GGCAATTAT	GTCTGACTT	GGCGACGCTT	TGGACGAACT		
501	GAACGCTCCC	TGCCATTG	GACAGGAT	CSTCCCGAA	GGCGTCGAG		
551	CCCAATACGA	CTGGCTGAT	GACTGGCGGC	GCTACGGCGC	AAAACCGCG		
601	TGAAACCAAT	CCCCCGAGCA	CACCAAGACC	CTGGCGGCGA	TAAGCGGGA		
651	ATGTCGGGGG	TTTACACAC	CCGAAATACG	GCTCAACCGC	CCCGTGCGTC		
701	TGTCCTCATCC	GGGTGATCC	CTCTACATCC	CCCGGAAGA	AAACACGTC		
751	TTGTCATCC	GGGGGGGACCA	AATCGAAAGC	GAAGGCCAG	CCCCGGCCAG		
801	CGTGGCTTCA	GGGTGGGAC	TCTCTCCCG	ACTCTATGCC	ATTCACCCG		
851	CCTTCGGG	AGCGGACATC	CTCGAAATCG	CCACCGGCC	GGCCCCACG		
901	CTCAACCA	ACAACCCCGA	ANTCCGTTAC	AACCGCGCC	GACSCCTGT		
951	TGAATC	GGCTTTTCC	GCAAGGTTT	CATGATCTCC	CCCGCGTAA		
1001	CCGGACGGG	CGCCGAGTTG	GCAGTGGCAC	TGTTTGACG	AAAAGACCGG		
1051	CCCGAACCGC	ATAAGAAG	CGGTTTGGC	TATATCCGA	GACAAGGATTA		
1101	A						

This corresponds to the amino acid sequence <SEQ ID 812; ORF126-1>:

```

1 MTRIAILGGG LSGRLTALQL AEQGYOIALF DKGCRGEHA AAYVAAAMLA
51 PAAEAEVATP EVVRLGRQSI PLWNRGIRCRRL NTHTMMQENG SLIVWHGQDK
101 FLSNEEVVRHL KRGGVADDEI VRWRADDIAE REPOLGGRFS DGIYLPTEQ
151 LDGRQILSAL ADALDELNPV CHWEHECVPF GLQAQYDWLI DCRGYGAKTA
201 WNGSPHENTST LRGIRGEVAR VYTPEITLNR PVRLLHPRYP LYIAPKENHV
251 FVIGATQIES ESQLAPASVRS GLELLSALYA IHPAFGEADI LEIATGLRPT
301 LNHNHFPIRY NMRKLLIEIN GLFRHGFMIS PAUTARAAARL AVALFDGKD
351 PERDKESCLA YIRQD*

```

10 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF126 shows 90.0% identity over a 180aa overlap with an ORF (ORF126a) from strain A of *N.meningitidis*:

	10	20	30	40	50	60
15 orf126.pep	MTRIAILGGGLSGRLTALQLAEQGYOIALFDKSCRGEHA	AAYVAAAMLA	PAAAXTVEATP			
orf126a	MTRIAILGGGLSGRLTALQLAEQGYOIALFDKRCRGEHA	AAYVAAAMLA	PAAEAEVATP			
	10	20	30	40	50	60
20 orf126.pep	70 80 90 100 110 120					
orf126a	EVVRLGRQSIPLWLRGIRCRRLNTTMQENGSLIVWHQDKPLSSEFVRHLKRGGVXTDII					
	70	80	90	100	110	120
25 orf126.pep	EVVRLGRQXPLWLRGIRCHLKTPLFAMMXENGSLIVWHQDKPLSNEFVRHLKRGGVADXI					
orf126a	VRWRADDIAEREPLQLGGRFSQDGYIPLTEQGLDGRQILSALADALDELNVPC					
	130	140	150	160	170	180
30 orf126.pep	VRWRADDIAEREPLQLGGRFSQDGYIPLTEQGLDGRQILSALADALDELNVPC					
orf126a	130 140 150 160 170 180					

The complete length ORF126a nucleotide sequence <SEQ ID 813> is:

```

1 ATGACCCGTA TCGCCATCTC CGGGCGCGGC CTCTCNGGAA GGCTGACCGC
51 ACTGAGCTT CGAACACAAAG GTTATCACAG TGCACTTTTC GATAAAGGCT
101 GCGCCGGGGG CGAACAGCGC GCGCCGCTATTC TTGGCGCGGC CATGCTCCGG
151 CTCGGCGGGG AGACCGCGGA AGCCACGCCG GRAGTGTCA GCGCTGGGAG
201 GCAGANCATC CGCTGTTCGGC GCGCGATCTCG ATGGCCATCTG AAAACGCCCTG
251 CCATGATGCA NGAAAAGCCG AGCCGTATTG TGTCGACCGG CGACGACAAA
301 CCTTATATCCA ACAGGATCTCG CGCCGATCTC AAACAGCGGGG CGCTAGCGGA
351 TGACNAATC GTCCGTTGGC CGCCGAGCGA CATCGCCGA CGGGAACCGC
401 AACCTGGGGG ACAGTTTCA GACGGCAATTC ACCCTGGCGAC CGAAAGGCCAG
451 CTGGACGGGG GCGCAATATCA GTCTGCACTT GCGGACCCCTT TGACGAACT
501 GAACATGGGG TGCCATGGG AACAGGAATTC TGCCCCCGAA GACTTGCAAG
551 CCCAATACGA CTGGCTGATC GACTGGCGGC GCTACGGCC AAAACCCGG
601 TGGACCATCC CCCCGGNNAA NACCAGCAC TGCGCGGGCA TACCGGGGCA
651 ATGGGGGGG GTTACACAC CGGAATACAC GCTAACCGGC CGCGTGGGCC
701 TGCTACACCC GCGCTATCCG CTNTACATCG CCCCCAAGGA AAACCNCGTC
751 TTGCTCATCG GCGGACCCCA AATGAAAGC GAAAGCCAAG CACCTGCCAG
801 CGTGCCTTCG GGGCTGGAAC TCTTATCCGG ACTCTATGCCGCTCCACCCCG
851 CCTTCGGGA AGCCGACATC CTGGAAATCG CCACCGGSCCT CGGCCACAG
901 CTCAAATCACC ACAACCCCGA ATTCCTTAC AATCCGCGCC GACCGCTGAT
951 TGAAATCACAC GCGCTTTCG GCGACCGTTT CATGATCTCC CGCCCGTAA
1001 CGCCCGCCCG CTCAGATTCG GCAGTGCGAC TGTTCGACGG AAAAGANGCG
1051 CGCGAACCGG ATGAAAGAAG CGGTTGGCC TATATCCGAA GACAAGATTA
1101 A

```

55 This encodes a protein having amino acid sequence <SEQ ID 814>:

```

1 MTRIAILGGG LSGRLTALQL AEQGYOIALF DKGCRGEHA AAYVAAAMLA
51 PAAEAEVATP EVVRLGRQXPLWLRGIRCRRL KTFAMMXENG SLIVWHGQDK
101 FLSNEEVVRHL KRGGVADDEI VRWRADDIAE REPOLGGRFS DGIYLPTEQ
151 LDGRQILSAL ADALDELNPV CHWEHECAFE DLQAQYDWLI DCRGYGAKTA
201 WNGSPHENTST LRGIRGEVAR VYTPEITLNR PVRLLHPRYP LYIAPKENXV

```

251 FVIGATQIES ESEQAPASVRS GLELLSALYA VHPAFGEADI LEIATGLRPT
 251 LNHNPETRY NRARRLIEIN GLFRHGMIS PAVTAAAVRL AVALFDGKXA
 351 PERDEEESGLA YIRRQD*

ORF126a and ORF126-1 show 95.4% identity in 366 aa overlap:

		10	20	30	40	50	60	
5	orf126a.pep	MTRIAILGGGLSGRLTALQLAEQQYQIALFDKSCRRGEHAAAYVAAMLA PAAAEAVEATP						
	orf126-1	MTRIAILGGGLSGRLTALQLAEQQYQIALFDKSCRRGEHAAAYVAAMLA PAAAEAVEATP						
10		10	20	30	40	50	60	
	orf126a.pep	EVVRIGRQXISIPLWRCIRCLNTHMMCENGSLIVWHGQDKPLSNEFVRHLKRGVADDXI	70	80	90	100	110	120
15	orf126-1	EVVRIGRQXISIPLWRCIRCLNTHMMCENGSLIVWHGQDKPLSNEFVRHLKRGVADDEI	70	80	90	100	110	120
		130	140	150	160	170	180	
20	orf126a.pep	VRWRADDIAEREPQLGGRFSDGIYLPTEGQLDGRQILSALADALDELNVPC HWEHECAPE	130	140	150	160	170	180
	orf126-1	VRWRADDIAEREPQLGGRFSDGIYLPTEGQLDGRQILSALADALDELNVPC HWEHECVP	130	140	150	160	170	180
25	orf126a.pep	DHQAYQDWLIDCRGYGAKTAWNQSPXXTSLRGIRGEVARVYTPEITLN RPVRLHHPRYP	190	200	210	220	230	240
	orf126-1	DHQAYQDWLIDCRGYGAKTAWNQSPXXTSLRGIRGEVARVYTPEITLN RPVRLHHPRYP	190	200	210	220	230	240
30	orf126a.pep	LYIAPKENHVFVIGATQIESSEQAPASVRSGLELLSALYA VHPAFGEADILEIATGLRPT	250	260	270	280	290	300
	orf126-1	LYIAPKENHVFVIGATQIESSEQAPASVRSGLELLSALYA IHPAFGEADILEIATGLRPT	250	260	270	280	290	300
35	orf126a.pep	LNNHNPEIRYNRARRLIEINGLFRHGMISPAVTA AARLVALFDGKXAPERDEESGLA	310	320	330	340	350	360
	orf126-1	LNNHNPEIRYNRARRLIEINGLFRHGMISPAVTA AARLVALFDGKDAPERDEESGLA	310	320	330	340	350	360
40	orf126a.pep	YIRRQDX						
	orf126-1	YIRRQDX						
45								

Homology with a predicted ORF from *N.gonorrhoeae*

ORF126 shows 90% identity over a 180 aa overlap with a predicted ORF (ORF126ng) from *N.gonorrhoeae*:

50	orf126.pep	MTRIAILGGGLSGRLTALQLAEQQYQIALFDKSCRRGEHAAAYVAAMLA PAAAEAVEATP	60
	orf126ng	MTRIAVLGGGLSGRLTALQLAEQQYQIELFDKGTRQGEHAAAYVAAMLA PAAAEAVEATP	60
55	orf126.pep	EVVRIGRQXISIPLWRCIRCLNTHMMCENGSLIVWHGQDKPLS SEFVRHLKRGVADDXI	120
	orf126ng	EVVRIGRQXISIPLWRCIRCLNTHMMCENGSLIVWHGQDKPLS SEFVRHLKRGVADDEI	120
	orf126.pep	VRWRADDIAEREPQLGGRFSDGIYLPTEGQLDGRQL XSALADALDELNVPC HWEHECVP	180
	orf126ng	VRWRADDEIAEREPQLGGRFSDGIYLPTEGQLDGRQ ILSALADALDELNVPC HWEHECAPQ	180

- 60 An ORF126ng nucleotide sequence <SEQ ID 815> was predicted to encode a protein having amino acid sequence <SEQ ID 816>:

1 MTRIAVLGGG LSGRLTALQL AEQQYQIELF DKGTRQGEHA AAYVAAAMLA

-449-

```

51 PAAEAATP EVIRLGROSQ PLWRGICRCL NLITMMQMQG SLIWHHQQD8
101 PLSSEFVRHL QHGKVADDEE VRWRAAARL REPQLGGGRFS DLPYLPTEGC
151 LQDQLRSAL ADALDELNPV CHWEHQECAPQ DLQAYQDWDVI DCRGYQAKTA
201 WNGQEPEITL RLRGIRGEVR FTRKPSRSTA PCACCTRRAIS STPKRPTKTS
251 SSSPARSKA KAKPPPAVY GNWSNPRSGS STPPSAPKTS SKWPRLGRPT
301 LNHNHPPEIRY SERRLRILIN GLFRHGMIS PAVTAANAVRL AVALEDPGDKA
351 PERDESSGLA YIGROD*

```

Further work revealed the following gonococcal DNA sequence <SEQ ID 817>:

10	1	ATAGCCGCTA	TGGCGCCGT	CGGAGGCCGC	CTTGGCGGAA	GCGTACGCCG
	51	ATAGGACATT	CGAACAGCAC	JTTATCATAG	TAACCTTCCT	GACAAAAGCC
	101	CCCGCCNAGG	CGAACACGCC	GCGCGCTATG	TTGGCCGCCG	AGATGTCGCA
	151	CTCTGGCGGG	AGGGCTGCGG	GCGAACAGCCG	GAAGTCATCA	GGCTGGGCGAG
	201	GCAGACAGCT	CCTGGTTGCG	GGCGCATCGC	ATGGCGCTTG	AACACGGCTCA
	251	CGATGATGAA	GGAAAAGGGG	AGCGCCTGATG	TGTGGCGCGG	GCGGACAGCA
	301	CCATATTCCA	GGAGGATCTG	CGGCCATCTG	AAAGCGGCCG	GGCTAGGGCAA
	351	GGCTGGGCG	GTCTGGTGGG	GGCGCGATGA	ATATGGCGGAA	GGAGGACGCC
	401	AACTCGGGCG	ACGTTTTCGA	GAAGCCGACT	ACCTGGCGAC	GAAGGGCGAC
	451	CTCGAGGGGG	GGCAAAATAT	GTCTGCACCT	GGCGACGCTT	GGTGCAGACT
	501	GAACGGCTCC	TGGCATTTGG	AAACAGCAATG	CCCCCCCCAT	GACTCTGAAGA
	551	CCCAATTAAGCA	CTGGTTGATC	GACTGCGCGG	GTACAGCGG	GAACAGCGCC
	601	TGGAAACCAT	CCCGGGAGGA	CACCGACGCC	TTGGCGGCCA	TACGGGGCGA
	651	ATGTGGGGGG	GTTTACACG	CCRAAATAC	GTCTAACCGC	CCCGTGGCGA
	701	TGTCTGGCCC	GGGTATCCGG	CTCTACATCG	CCCGNNAGAA	AAACCAAGCG
	751	TTCTGCATCG	GGGGCGGACCA	AATCGAAAGGC	GAAGACCGAA	CCCCGGCGCA
	801	CGTACCTTCG	GGGGCTGGAA	TCTTATCGGC	GTCTATGATC	GTCCACCCGGG
	851	CTCTGGGGGA	AGCGCGATAC	CTGGAATTCG	CGGCCGCGCT	GGGCCCGGAC
	901	CTTAACCAACCA	ACAAACCGGA	AAATCGCTAC	AGCGCGGAA	GGCGCCCTCAT
	951	CGGAATTAAC	GGGGCTTTCCTT	GGCGACGCGT	TATGATTTCC	CCGGCGCTTA
	1001	CGGGCGGCCG	CGTCAGATTC	CGAGTCGGAC	TTGTTGACCG	GGAGAACGCG
30	1051	CGCGGAACCG	ATGAGAAGRAA	CGGTTTGGCG	TATATCGGAA	GACNAGAGTTA
	1101	B-				

This corresponds to the amino acid sequence <SEO ID 818: ORF126ng-1>

1 MTRIAVLLGG SLSRLTQLL AEGQYQIEFL DGKTRGQEAE AAVAAMLA
 51 PAAEAEATPE EVIRLGRQSI FLWLRGCRCL NTLLTMWQNEA SLIVWHHQKQD
 101 PLSSVEFVRH KRGGVADDE VRMRADAEZAS REPOLGRQSL DGYLTPEVG
 151 LDGRQLSAL ADALDELNPV CHWEHECAPO DLQAYDWNV1 DCRGYQAKTA
 201 WQNQSPHTES LRIGIRGEVAR VPTTEIILNR PVRLHLPPRY LYIAPKENH
 251 FVIGATOES ESQAPAVSLES GLELLSALYA WHPFAFEID LELIAAGLRPT
 301 LNHNHFPEE SRERHLIEIN GLFRHGFMIG_PAVTAJAVRL AVALFDGKD
 351 PERDESGLA YIGRD+
 40

ORF126ng-1 and ORF126-1 show 95.1% identity in 366 aa overlap:

-450-

		250	260	270	280	290	300
	orf126-1.pep	LYIAPKENHVFIGATQIESESQAPASVRSGLELLSALYAHIPAFGEADILEIAGLRPT					
5	orf126ng-1	LYIAPKENHVFIGATQIESESQAPASVRSGLELLSALYAHIPAFGEADILEIAAGLRT					
		250	260	270	280	290	300
		310	320	330	340	350	360
10	orf126-1.pep	LNHHNPEIRYNARRRLIEINGLFRHGMISPAVTAARLAVALFDGKDAPERDKESSGLA					
	orf126ng-1	LNHHNPEIRYSRERRRLIEINGLFRHGMISPAVTAAVRLAVALFDGKDAPERDEESGLA					
		310	320	330	340	350	360
15	orf126-1.pep	YIRRQDX					
	orf126ng-1						
		YIGRQDX					

Furthermore, ORF126ng-1 shows homology to a putative *Rhizobium* oxidase flavoprotein:

```

gi|2627327 (AF004408) putative amino acid oxidase flavoprotein [Rhizobium etli]
Length = 327
Score = 169 bits (423), Expect = 3e-41
Identities = 112/329 (34%), Positives = 163/329 (49%), Gaps = 25/329 (7%)
Query: 3 RIAVGGGLSRLTAQLAEQGYQIELFDKGTQGEHXXXXXXXXXXXXXXXXXXXXX 62
      RI V G +G+ A Q L G+++ L ++ G
25 Sbjct: 2 RILVNAGVAGLTVAWQLYRHGFRVTLAERAGTVGA-GASGFAGGMLAPWCERASEEPV 60
Query: 63 IRLGRQSITPLWRGTRCRNLTTMQEONGSLVWHDQDKPLSSEFVRHLRKRGVADDEIVR 122
      + LGR + W + G-L+V G-D F R G DE+
      Sbjct: 61 LTLCRLAADMWEAA----LPGHVHRRGTLVLAGGDRDTGELDRFSRRTS-GWEWLDEVA- 113
Query: 123 WRADEAIREPDLGRFSDGYIYLPTEGLDGRQILSALADALDELNVPCHEHECAPQDL 182
      IA EP L GRF ++ E LD RQ L+ALA L++ + +
      Sbjct: 114 -----IAALEPDLAGRFRRALFFRQEALHDPQLALAALAGLEDAMRLTLLG---VVGES 165
Query: 183 QAQYDWDWIDCRGYGAKTAWNQSPHTSTLRGIRGEVARVYTPEITLNRPVRLLHPRYPLY 242
      +D V+DC G LRG+RGE- V T E+L+RPVRLLHPR+PY
      Sbjct: 166 DVDRHDRVVDCTGAA-----QIGRLPGLRGVRGEMLCVETTEVSLSRVRLLHPHPHY 218
Query: 243 IAPKENHVFIGATQIESESQAPASVRSGLELLSALYAHIPAFGEADILEIAAGLRTIN 302
      I P++ F+GAT IES- P + RS +ELL+A YA-HPAFGEA + E AG+RP
      Sbjct: 219 IVPRDKNRMFMGATMIESDDGGPITARSIMLAAAAMHPAFGEARVTTETGAGVRPAYP 278
Query: 303 HHHNPEIRYSRERRRLIEINGLFRHGMISP 331
      + P R ++E R + +NGL+RHC+-++P
45 Sbjct: 279 DNLP---RVTQEGRTLHVNGLYRHCGFLAP 305

```

This analysis suggests that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 97

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID

50 819>:

```

1 ATGACTGATA ATCGGGGGTT TACGCTGGTT GAATTAATAT CAGTGGTCTT
51 GATATTGTC TGTACTTGCTT TAMTTGTTTA TCCGGAGCTAT CGCAATTATG
101 TTGAGAAAGC AGAACATAAT GCAGTCGCC CAGCCTGTT AGAAAATGCA
151 CATTATGGA AAAAGTTTA TCTGCAGAAT GGGAGGTTTA AACAAACATC
201 TACCAAGTGG CCAAGTTGC CGATTAAGA GGCAGAGGCC TTTTGATATC
251 GTTGAATGG AATCG+CGCG CCGG...GCTT TAGACAGTAA ATTCTATGTTG
301 AAAGCCGTAG CCATAGATAA AGATAAAAAT CCTTTTATTA TTAAAGATGAA
351 TGAAATCTA GTAAACCTTA ATTTGCAAGA AGTCGGCGAG TTCTGTTGAT
401 GACGGGCTGG ATTATTTAA AGGAAATGAT AAGGACTGCA AGTACTTAA
451 GTAG

```

This corresponds to the amino acid sequence <SEQ ID 820; ORF127>:

```

1 MTDNRGFTLV ELISVVLILS VLALIVYPSY RNYVEKAKIN AVRAALLENA
51 HFMEKFYLNQ GRFKQTSTKW PSLPIKEAEG FCIRLNGIWA RXALDSKFLM
101 KAVAIKDKNP PTIICKMNEL VTFICKKSAS CSDGLDYFK GNDKDCKLK
151 *

```

Further work revealed the following DNA sequence <SEQ ID 821>:

```

1 ATGACTGATA ATCGGGGGTT TAGCCTGGTT GAATTAATAT CAGTGGTCTT
51 GATATTGCT GTACTTGCT TAATTTGTTA TCCGAGCTAT CGCAATTATG
101 TTGAGAAAGC AAAGATAATA GCAGTCGGG CAGCCTTGT AGAAAATGCA
151 CATTTTATGG AAAAGTTTA TCCTGCAGAAT GGAGGATTAA AAAAAACATC
201 TACCAAGTGG CCAAGTTGCG CGATTAAGAAG GCGAGAAGG TTGTTGATCC
251 GTTGAATGG AATCGGCCG GGGGCTTTAG ACAGTAAATT CATGTTGAAG
301 GCGTAGGCCA TAGATAAAGA TAAAAATCTT TTATTATTA AGATGAATGA
351 AAATCTAGTA ACCTTTATT GCAGAAGTC CGCCAGTTG TGTTAGTGACG
401 GGCTGGATTA TTTAAAGGA ATGTATAAGG ACTGCAAGTT ACTTAAGTAG

```

This corresponds to the amino acid sequence <SEQ ID 822; ORF127-1>:

```

1 MTDNRGFTLV ELISVVLILS VLALIVYPSY RNYVEKAKIN AVRAALLENA
51 HFMEKFYLNQ GRFKQTSTKW PSLPIKEAEG FCIRLNGIAR GALDSKFLM
101 AVAIKDKNP PTIICKMNEL VTFICKKSASS CSDGLDYFKG NDRDCKLLK*

```

20 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N. meningitidis* (strain A)

ORF127 shows 98.0% identity over a 150aa overlap with an ORF (ORF127a) from strain A of *N. meningitidis*:

	10	20	30	40	50	60
25 orf127.pep	MTDNRGFTL V LISVVLILS V LALIVYPSY R NYVEKAKIN A VRAALLENA H FMEKFYLN					
orf127a	MTDNRGFTL V LISVVLILS V LALIVYPSY R NYVEKAKIN A VRAALLENA H FMEKFYLN	10	20	30	40	50
		10	20	30	40	60
30 orf127.pep	GRFKQTSTKWPSPPIKEAEGFCIRLNGI V ARALDSKFLK V AAIKDKNPFIICKMNEL	70	80	90	100	110
orf127a	GRFKQTSTKWPSPPIKEAEGFCIRLNGI-ARGALDSKFLK V AAIKDKNPFIICKMNEL	70	80	90	100	110
		70	80	90	100	120
35 orf127.pep	VTFICKKSASSCSGDLDYFKGN D CKKLKX	130	140	150		
orf127a	VTFICKKSASSCSGDLDYFKGN D CKKLKX	120	130	140	150	

The complete length ORF127a nucleotide sequence <SEQ ID 823> is:

```

1 ATGACTGATA ATCGGGGGTT TAGCCTGGTT GAATTAATAT CAGTGGTCTT
51 GATATTGCT GTACTTGCT TAATTTGTTA TCCGAGCTAT CGCAATTATG
101 TTGAGAAAGC AAAGATAATA GCAGTCGGG CAGCCTTGT AGAAAATGCA
151 CATTTTATGG AAAAGTTTA TCCTGCAGAAT GGAGGATTAA AAAAAACATC
201 TACCAAGTGG CCAAGTTGCG CGATTAAGAAG GCGAGAAGG TTGTTGATCC
251 GTTGAATGG AATCGGCCG GGGGCTTTAG ACAGTAAATT CATGTTGAAG
301 GCGTAGGCCA TAGATAAAGA TAAAAATCTT TTATTATTA AGATGAATGA
351 AAATCTAGTA ACCTTTATT GCAGAAGTC CGCCAGTTG TGTTAGTGACG
401 GGCTGGATTA TTTAAAGGA ATGTATAAGG ACTGCAAGTT ACTTAAGTAG

```

This encodes a protein having amino acid sequence <SEQ ID 824>:

```

1 MTDNRGFTLV ELISVVLILS VLALIVYPSY RNYVEKAKIN TVRAALLENA
51 HFMEKFYLNQ GRFKQTSTKW PSLPIKEAEG FCIRLNGIAR GALDSKFLM
101 AVAIKDKNP PTIICKMNEL VTFICKKSASS CSDGLDYFKG NDRDCKLLK*

```

-452-

ORF127a and ORF127-1 show 99.3% identity in 149 aa overlap:

		10	20	30	40	50	60
	orf127a.pep	MTDNRGFTLVELISVVLILSVALALIVVPSYRNYVEKAKINTVRAALLENAHFMEKFYLQN					
5	orf127-1	MTDNRGFTLVELISVVLILSVALALIVVPSYRNYVEKAKINAVRAALLENAHFMEKFYLQN	10	20	30	40	50
	orf127a.pep						
10	orf127-1						
	orf127a.pep	GRFKQTSTKWPSPSLPIKEAEGFCIRLNGIARGALDSKFMKAVAIKDKNPFIKKMNEVLV	70	80	90	100	110
	orf127-1	GRFKQTSTKWPSPSLPIKEAEGFCIRLNGIARGALDSKFMKAVAIKDKNPFIKKMNEVLV	70	80	90	100	110
	orf127a.pep						
15	orf127-1						
	orf127a.pep	TFICKKSASSCSGDLDFKGNDKDCKLKX	130	140	150		
	orf127-1	TFICKKSASSCSGDLDFKGNDKDCKLKX	130	140	150		

20 Homology with a predicted ORF from *N.gonorrhoeae*

ORF127 shows 97.3% identity over a 150 aa overlap with a predicted ORF (ORF127ng) from *N.gonorrhoeae*:

	orf127.pep	MTDNRGFTLVELISVVLILSVALALIVVPSYRNYVEKAKINAVRAALLENAHFMEKFYLQN	60
25	orf127ng	MTDNRGFTLVELISVVLILSVALALIVVPSYRNYVEKAKINAVRAALLENAHFMEKFYLQN	60
	orf127.pep	GRFKQTSTKWPSPSLPIKEAEGFCIRLNGIARVALDSKFMKAVAIKDKNPFIKKMNEVLV	120
	orf127ng	GRFKQTSTKWPSPSLPIKEAEGFCIRLNGI-ARGALDSKFMKAVAIKDKNPFIKKMNEVLV	119
30	orf127.pep	TFICKKSASSCSGDLDFKGNDKDCKLKX	150
	orf127ng	TFICKKSASSCSGDLDFKGNDKDCKLKX	149

The complete length ORF127ng nucleotide sequence <SEQ ID 825> is:

35	1	ATGACTGATA ATCGGGGGTT TACACTGGGT GAATTAAATAT CAGTGGCTT	
	51	GATATTGCT GTACTTGCT TAATTGTTTA TCCGAGCTAT CCCAATTATG	
	101	TGGAGAAAGC AAAGATAAAT GCAGTCGGGG CGACGTGTTT AGAAAATGCA	
	151	CATTTTATGG AAAAGTTTA TCTGCAGAAT GGGAGATTTA AACAAACAT'C	
40	201	TACCAAATGG CCAAGTTTG CGATTAAGAA GGCGAGAAGG TTITGTATCC	
	251	GTTGAATGG AA'TCGGGGGG GGGGCCTTGA ACAGTAANTT CATGTTGAAG	
	301	GGGTAGGCC TAGATAAGA TAAAAATCTT TTATTATTA AGATGAAATGA	
	351	AAACTCTAGTA ACCTTTATTC CGAACAGAATC CGGCAGATTCC TGAGTGAAG	
	401	GGCTGGATTA TTTAAAGGG ATTGATAAGG ACTGCAAGTT ACTTAACTGAG	

This encodes a protein having amino acid sequence <SEQ ID 826>:

45	1	MTDNRGFTLV ELISVVLILS VLALIVVPSY RNYVEKAKIN AVRAAFLENA	
	51	HFMEKFYLQN GRFKQTSTKWPSPSLPIKEAEGFCIRLNGIARVALDSKFMKAVAIKDKNPFIKKMNEVL*	
	101	AVAIKDKNPFIKKMNEVLV TFICKKSASSCSGDLDFKGNDKDCKLK*	

ORF127ng and ORF127-1 show 100.0% identity in 149 aa overlap:

50	orf127-1.pep	MTDNRGFTLVELISVVLILSVALALIVVPSYRNYVEKAKINAVRAALLENAHFMEKFYLQN	10	20	30	40	50	60
	orf127ng-1	MTDNRGFTLVELISVVLILSVALALIVVPSYRNYVEKAKINAVRAALLENAHFMEKFYLQN	10	20	30	40	50	60
55	orf127-1.pep							
	orf127ng-1							
	orf127-1.pep	GRFKQTSTKWPSPSLPIKEAEGFCIRLNGIARGALDSKFMKAVAIKDKNPFIKKMNEVLV	70	80	90	100	110	120
	orf127ng-1	GRFKQTSTKWPSPSLPIKEAEGFCIRLNGIARGALDSKFMKAVAIKDKNPFIKKMNEVLV	70	80	90	100	110	120

	130	140	150
orf127-1.pep	TFICKKSASSCSGDLDFKGNDKCKLLKX		
orf127ng-1	TFICKKSASSCSGDLDFKGNDKCKLLKX		
	130	140	150

5

This analysis, including the fact that the predicted transmembrane domain is shared by the meningococcal and gonococcal proteins, suggests that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

10 Example 98

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 827>

1	..GTGTCGCTGG	CTTCGGTGT	TGCTCTCAA	ATCTCCCTTT	ACGAAGATT
51	CAACAAATG	CGGAAAACCC	GTTGAGCAT	CTGGGGTTTT	CTTGTCCAAAT
101	ATTATCTGG	GGTTTACAGCA	GGGGTTATTTC	GATTGGATG	CGGACGAGAAA
151	CCCGGTACTG	CATACTGGT	CTTGGCAGGT	AGAGGAAACAG	TATTACCTCC
201	TGATTCCTCT	TTTCTGATA	TTTCTGCTCA	AAAAAAACCA	ATGCCAACGG
251	GTGCTGCCTA	ACATCACAT	CATCCCTGTTT	TTGATTTGA	CTGCCCTCATC
301	GTTTTGCGCA	ACGGGGTATT	ATACCAGCAT	CCTCAACCAA	CCAAATACTT
351	ATTACCTTTC	GACACTGAGG	TTTCCGGAGG	TGTTGGCAAGG	TTGGCTGCTG
401	GCCTTITACG	GGCRAACCGA	AAAAGGCGAG	CGSGAAACAG	CAAAATGGAAA
451	ACGGCAGTTG	CTTCTCACTC	TCTGCTTGGG	CGCATTTGCTT	GGCTGCTCTG
501	TCTGATTTGA	AAACAAACAT	CGCTTTTACAT	CGGAAATGAC	CCTGCTCTT
551	CCCTGGCTGC	TGAGGGCACT	GCTTATCCGG	AGTATGCAAT	ACGGGACACT
601	TCGGACCCGC	ATTCCTGCG	CAAGGCCCCAT	CGTATTGTC	GGCAAATATC
651	CTTATTCCTC	ATACATGTC	ATCCTGGATT	TTATTGCTTT	CGCTCCGCTC
701	ATTAGGAGCCG	GGAAACAGCT	CGGACTGCC	CGCCG..	

This corresponds to the amino acid sequence <SEQ ID 828; ORF128>:

1	..VSLASVIASQ	IFLYEDFNQM	RKTVELSAVF	LSNIYLGFQQ	GYFDLSADEN
51	PVLHIIISLAV	EEQVYLLYPL	LLIFCKKRTS	SLRVLRNISI	IILFLILTASS
101	FLPSGFYPDFI	LQNQNTTYLIS	TLFPEFLLAG	SLLAVYQQTQ	NGRRQTAANGK
151	RQLLSSLCFG	ALLACLEVID	KHNPFPIGM	LLLPCLLTAL	LIRSMQYGL
201	PTRLSASP1	VFVGKISYSL	YLYHWIFIAF	APLIRGGKQL	GLPA..

Further work revealed the complete nucleotide sequence <SEQ ID 829>:

1	ATGCAAGCTG	TCCGATACAG	ACCGGAAATT	GACCGATTGC	GGGGCGCTGC
51	CGTGCTATCC	GTCATGATT	TCCACTGAA	TAACCGCTGG	CTGCCCCGGAG
101	GATTCTCGG	GTTGAGCAT	TTCCTTGCA	TCTCAAGATT	CCTCATTAAC
151	GGGATCACTC	TTTCGAATTC	ACAGAACCGT	TCTTTTTCTT	TCGGGGATT
201	TTATACCCGC	AGGATTAAGC	GGATTATTC	TGCCCTTAT	GCGCCGCTGT
251	CGCTGGCTTC	GAATGGTGC	TCCAAATCT	TCTCTTACGA	AGATTTCAAC
301	CAATGGCGA	AAACGGTGG	GCTTCTCGG	TTTTCTTGT	CCATATTAA
351	TCTGGGGTT	CAGGGGGGT	ATTCGGATTG	GAGTGGCCGC	GAGAACCCCG
401	TACTGCATAT	CTGGCTTTC	CGCAGTAGGG	AACAGTATT	CCTCTGTAT
451	CCCTTTCG	TGATATTTC	CTGCAAAAAAA	ACCAATTCG	TACGGGTGCT
501	GGCTAACATC	AGCATCATC	TGTTTTGAT	TTGACTGCC	TCATCGTTT
551	TGCGAACAGGG	GTTCATATAC	GACATCTCCA	ACCAACCCAA	TACTTATTAC
601	CTTTCGACAC	TGAGGTTTC	CGAGGTGTC	CGAGGTTC	TGCTGGCGGT
651	TTACGGGCAA	ACGCAAAACG	CGAGACGGCA	ACACGAAAT	GGAAAAACGGC
701	AGTTGCTTTC	ATCACTGTC	TTCGGCGAT	TGCTTGCTG	CTCTGCTGTG
751	ATTGACAAAC	AACTACCGT	TATCCCGGGA	ATGACCTCTG	TCTTCCCTG
801	CTCGCTGRCG	GGACTGCTTA	TCCGGGTAT	CGAATACGGG	ACACTCCCGA
851	CCCCCATCTC	TGCGCAAGC	CCCATCGTC	TTGTCGGCAA	AATCTCTTAT
901	TCCTTAAAC	TGTCACATC	GATTTTATT	GTTTCCCCC	ATCACATAC
951	AGGCCACAA	CAGCTGGAC	TGCGCTGGGT	ATCGGGGTT	GGCGCGTGTG
1001	CGGCCGATT	TTCCCTGTTG	AGTTTATT	TGATTTGAAACA	GGCGCTTAGA
1051	AAACGGAAAGA	TGACCTTCA	AAAGGCAATT	TTCTGCTCT	ATCTCGCCCC
1101	GTCCCTGATA	CTTGTGCGTT	ACAACTGTG	CGCAAAAGGGG	ATATTGAAC
1151	AGGAAACACCT	CGCGCGCTG	CCCGCGCGC	CCCTTGCTG	GGAAAATCAT

1201 TTTCGGAAA CCGTCTGAC CCTCGCGCAG TCGCACGCCG GACACCTGAG
 1251 GGGTTTCG GATTATGTCG CGACCGCGA AGGGTGAA GCCAAATCC
 1301 TGCCCTCGA TTGGAGSTGT TGGTTTGGG TAGATGAGAA GTGCGCAC
 1351 AACCGTTAT GTCGAAATA CCGGGATGAA GTTGAARAG CGAACGCCG
 1401 TTTCATGCC CAACTATG ATTTCAGGAT GGGCGGCCAG CCTSITGCCA
 1451 GATTGGAGC GCAATCTTC CTAATACCG GGTTCCAGC CGATTCAGG
 1501 GAAACCTCA AAAGGATAGC CCGCCTCAA CGCGCTATG TTTTTGCAAA
 1551 CAACACATCA ATACGCGCTT CGCCCGCTGAG GGAGGAAAGAT TTGAAAAGAT
 1601 TTGGCCAAA CAAATATCT CGCCCGATTC AGGCTATGGA CGACATCGC
 1651 AAAGACNATC AGCCGCTTC TGATTTGATT AAAGATATTIC CCAATGTCGA
 1701 TTGGGTGCA GCACAAAAAA ACCTGCCCCA AAACACGGTC GAATATACG
 1751 GCGCTATCT TTACGGCAG CAAGACCCAG TGACCTATTAT CGGTTCTTAT
 1801 TATATGGGC GGAATTCCA CAAACAGAA CGCCGTCTTA AATCTTCCA
 1851 CGCGCGGCGA TTGCAAGTAG

15 This corresponds to the amino acid sequence <SEQ ID 830; ORF128-1>:

1 MQAVRYPPEI DGLRAVAVLS VMIFHILNNRW LPGGFLGVDI FVISGFLIT
 5 GITLSEIQNG SFSFRDFYTR RIKRIPFAI AAVSLASVIA SQFLYEDFN
 10 QMRKTQVNL VFLSNIYLGF QQGYFDLSDA ENPVHLHISL AVEEQYLLY
 15 PLLIFCCKRL TSILRVLNRI SIIFLFLITA SSFLPSGFYT DIAQNPNITY
 20 LSTLRFPELL AGSLALAVYQD TONGRQQTAN GKRQLLSSLC FGALLACLFV
 25 IDKHNPTPEL MTLPLLPCP ALLIRNSMQYD TLPTPLRSAC PIVFVGKISY
 30 SLIYHWIFI AFAHYITGDR QLGPAVASV AALTAGFSLL SYLIEQPLR
 35 KRMKTFKAF FCFLYAPSLI LVGVNYIARG ILRKQEHIRPL PGAPLAEEHH
 40 FPEVTLTLD SHAGHLRGFL DYVGSREGWIK AKILSLDSEC LWVVDKEKLAD
 45 KSNQAEVLDI VERAKEAVFIA QFYDLRNGQ PVPFRFAQOSF LIPGPAPRFR
 50 ETVKRIAALK PVVYFANNNT ISRSPLBEEK LKRFANQYIL RP1QAMGDIG
 55 KSNQAEVLDI KDIPIPNHVWD AKQYLPKNTV EIYGRYLYGD QDHLTYFGSY
 60 YMGRFHKKHE RLLKSSHHGA LQ*

Computer analysis of this amino acid sequence gave the following results:

30 Homology with hypothetical integral membrane protein HI0392 of *H.influenzae* (accession number U32723)

ORF128 and HI0392 show 52% aa identity in 180aa overlap:

Orf128: 1	VSLASVIASQIFLYEDNFQMRRKTVELSAVFLSNIYLGFQGYFDLSADENPVLHILNSLAV	60
	++L S IAS IF+Y DN++RRT+EL + FLSN YLG QQYFDLSA+ENPVHLHISLAV	
35 HI0392: 46	MAVLSFVIASIAFIYNDNTKLRKTIELIAFLSNFYLGLTQGYFDLSANENPVLHILNSLAV	105
Orf128: 61	EQXXXXXXXKXIFCCKRLTSILRVLNRI SIIFLFLITA SIIFLFLITASSFLPSGFYT DILNQPNITYL	120
	I KK + ++VL I++ILF IL A+SF+ * FV ++L+QPN YIYL	
40 HI0392: 106	EQQQYLIFPLLILAYKKFREVKVLFITLILFLILLATSFSVANFYKEVLHQPNIYLLS	165
Orf128: 121	TLRFPPELAGSLLAVYGTQNTGRRQTAANGRQLLSSLCFGALLACLFVIDKHNPFIGPMG	180
	LRFPPELL GSLLA-Y N + Q + +L+ L L +CLF+++ + FTIGP+T	
45 HI0392: 166	NLRFPPELVGSLLAIYHNLSN-KVQLSKQVNNIILALSTLLFCFLIMNNNNIAPGIGT	224

Homology with a predicted ORF from *N.meningitidis* (strain A)

45 ORF128 shows 98.0% identity over a 244aa overlap with an ORF (ORF128a) from strain A of *N.meningitidis*:

orf128.pep	10 20 30
	VSLASVIASQIFLYEDNFQMRRKTVELSAV
50 orf128a	
	ILSEIQNGSFSFRDFYTRRIKRYPAFIAVSLASVIASQIFLYEDNFQMRRKTVELSAV
	60 70 80 90 100 110
55 orf128.pep	40 50 60 70 80 90
	LSNIYLGFGQGYFDLSADENPVLHILNSLAVEEQYIYLXPLLLLIFCKKTKTSLRVLNRISI
orf128a	
	LSNIYLGFGQGYFDLSADENPVLHILNSLAVEEQYIYLXPLLLLIFCKKTKTSLRVLNRISI
	120 130 140 150 160 170
60 orf128.pep	100 110 120 130 140 150
	TFLFLITASSFLPSGFYTIDINQPNITYLSTLRFPPELAGSLLAVYGTQNGRRQTAANG

-455-

	orf128a	IILFLILITATSPLSPGSFYDILNQPNTYLYSLTRLFPELLAGSILLAVYQGTQNQRQTANGK 180 190 200 210 220 230
5	orf128.pep	160 170 180 190 200 210 RQLLSSLCFVGALLACLVIFIDKHNPFFIPGMLTLLPCCLTALLIRSMOYGTLPTRILSASPIL
	orf128a	240 250 260 270 280 290 RQLLSSLCFVGALLACLVIFIDKHNPFFIPGMLTLLPCCLTALLIRSMOYGTLPTRILSASPIL
10	orf128.pep	220 230 240 VFVKGKISYSLYLYHWIFIARFAPLIRGGKQLGLP
	orf128a	300 310 320 330 340 350 VFVKGKISYSLYLYHWIFIARFAHYITYGDQKLGPLPAVASAVAALTAGFSLLSYLIEQPLRK
15	orf128a	360 370 380 390 400 410 KMFTKKAFFCCLYLSPPLLWGYNLYARGK1QEHRLPFGAPLAEEHNHFPTVTLGDSS

The complete length ORF128a nucleotide sequence <SEQ ID 831> is:

This encodes a protein having amino acid sequence <SEO ID 832>;

60	1 M ^Q A ^V R ^E P ^I D ^G L ^R A ^V A ^S V ^M H ^F L ^R N ^R L ^G E ^G P ^G V ^D F ^P V ^I S ^G L ^F L ^I T ^I 51 G ^I L ^I S ^E C ^I Q ^N S ^F R ^E P ^T V ^R R ^I K ^R IY ^P A ^F A ^V A ^S L ^V A ^S S ^O I ^E L ^I Y ^F P ^I S ^G 101 O ^N M ^R K ^T V ^S L ^S A ^S V ^P L ^N S ^I Y ^G Q ^O G ^V F ^D I ^S A ^D E ^N V ^L H ^I W ^S L ^T A ^V E ^N P ^T Y ^S 151 P ^L L ^I F ^C C ^K E ^I T ^K S ^L R ^V R ^I N ^R S ^I I ^L L ^F L ^I T ^S L ^T P ^L S ^G F ^T D ^I L ^O N ^P V ^T Y ^S 201 L ^S T ^L R ^A P ^E I ^S A ^G S ^L V ^A Y ^G T ^Q G ^R R ^T A ^N K ^G R ^L O ^S F ^G A ^L L ^A C ^I P ^E 251 I ^D K ^H N ^P F ^I P ^G M ^T L ^I F ^C L ^I T ^S A ^L L ^I R ^S Y ^M G ^I T ^L P ^R I ^S L ^T F ^I V ^G K ^I S ^S 301 S ^I L ^Y L ^W I ^I F ^A Z ^H Y ^T G ^I D ^G L ^C P ^A V ^S A ^L T ^G A ^T S ^I S ^Y Y ^L E ^P A ^N R ^I 351 R ^K R ^M T ^E K ^F A ^S F ^C L ^Y A ^L S ^I L ^U G ^V N ^Y R ^G L ^I K ^O E ^N R ^I P ^G A ^P L ^I E ^N R ^I 401 F ^P E ^T V ^T L ^G D ^I S ^A H ^G R ^L E ^I D ^Y G ^S R ^E C ^W A ^K I ^L S ^D L ^S E ^R M ^L W ^W V ^O K ^A C ^I 451 M ^P L ^C X ^Y R ^D U ^E K ^A R ^B A ^T D ^F Y ^O R ^M G ^G O ^I B ^R E ^A B ^S O ^I L ^I R ^E G ^B E ^A P ^I
65	

501 ETVKRKIAAVK PVVYFVANNTS ISRSPLREKK LKRFAANQYL RPIQAMGDIG
 551 KSNQAVFDLI KDIPVNHWVD AQQYLPKRTNV EIYGRYLYGD QDHLYTFGSY
 601 YNGREFHKHE RLLKSSRDGA LQ*

ORF128a and ORF128-1 show 99.5% identity in 622 aa overlap:

5	orf128a.pep	MQA VRYR PR EIDGL RAV AVAL SV MIF HLN RR N WLP GGFL GL GV D IFF V I S G F L I T G I I L L E I Q N G
	orf128-1	MQA VRYR PEI DGL RAV AVAL SV MIF HLN RR N WLP GGFL GL GV D IFF V I S G F L I T G I I L L E I Q N G
10	orf128a.pep	SFS FR DF YTR RI KI YPA F A V A V S L A S V I A S Q I F L Y E D F N Q M R K T V E L S A V F L S N I Y L G F
	orf128-1	SFS FR DF YTR RI KI YPA F A V A V S L A S V I A S Q I F L Y E D F N Q M R K T V E L S A V F L S N I Y L G F
15	orf128a.pep	Q Q G Y F D L S A D E N P V L H I W S L A V E E Q Y Y L Y P L L I F C C K K T K S L R V L R N I S I I L F L I L T A
	orf128-1	Q Q G Y F D L S A D E N P V L H I W S L A V E E Q Y Y L Y P L L I F C C K K T K S L R V L R N I S I I L F L I L T A
20	orf128a.pep	T S F L P S G F Y T D I L N Q P N T Y Y L S T L R F P E L L A G S L L A V Y Q G T Q N G R R Q T A N G K R Q L L S S L C
	orf128-1	S S F L P S G F Y T D I L N Q P N T Y Y L S T L R F P E L L A G S L L A V Y Q G T Q N G R R Q T A N G K R Q L L S S L C
25	orf128a.pep	F G A L L A C L E V I D K H N P F I P G M T I L L P C L L T A L L I R S M Q Y G T L P T R I L S A S P I F V G K I S Y
	orf128-1	F G A L L A C L E V I D K H N P F I P G M T I L L P C L L T A L L I R S M Q Y G T L P T R I L S A S P I F V G K I S Y
30	orf128a.pep	S L Y L Y H W H I F A F A H Y I T G D K Q L G L P A V S A V A L T A G F S L L S Y L I E Q P L R K R K M T F K K A F
	orf128-1	S L Y L Y H W H I F A F A H Y I T G D K Q L G L P A V S A V A L T A G F S L L S Y L I E Q P L R K R K M T F K K A F
35	orf128a.pep	F C L Y L A P S L I L V G Y N L Y A R G I L K O E H L R P L P G A P I A E N H P E T V L T L G D S H A G H L R G F L
	orf128-1	F C L Y L A P S L I L V G Y N L Y A R G I L K O E H L R P L P G A P I A E N H P E T V L T L G D S H A G H L R G F L
40	orf128a.pep	D Y V G S R E G W K A K I L S L D S E C L V W V D E K L A D N P L C R K Y R D E V E K A E A V F I A Q F Y D L R M G G Q
	orf128-1	D Y V G S R E G W K A K I L S L D S E C L V W V D E K L A D N P L C R K Y R D E V E K A E A V F I A Q F Y D L R M G G Q
45	orf128a.pep	P V P R F E A Q S F L I P G F P A R F R E T V K R I A A V K P V V F V A N N T S I S R S P I L R E E K L K R F A A N Q Y L
	orf128-1	P V P R F E A Q S F L I P G F P A R F R E T V K R I A A V K P V V F V A N N T S I S R S P I L R E E K L K R F A A N Q Y L
50	orf128a.pep	R P I Q A M G D I G K S N Q A V F D L I K D I P V N H W D A Q K Y L P K N T V E I Y G R Y L Y G D D H I T Y F G S Y
	orf128-1	R P I Q A M G D I G K S N Q A V F D L I K D I P V N H W D A Q K Y L P K N T V E I Y G R Y L Y G D D H I T Y F G S Y
55	orf128a.pep	Y N G R E F H K H E R L L K S S R D G A L Q X
	orf128-1	Y N G R E F H K H E R L L K S S H G G A L Q X

Homology with a predicted ORF from *N.gonorrhoeae*

50 ORF128 shows 93.4% identity over 244 aa overlap with a predicted ORF (ORF128ng) from *N. gonorrhoeae*:

	orf128.pep	V S L A S V I A S Q I F L Y E D F N Q M R K T V E L S A V F	30
55	orf128ng	I L S E I Q N G S F S F R D Y T R R I K R I Y P A F I A V A S L A S V I A S Q I F L Y E D F N Q M R K T I E L S T V F	112
	orf128.pep	L S N I Y L G F Q Q Y G F D L S A D E N P V L H I W S L A V E E Q Y Y L Y P L L I F C C K K T K S L R V L R N I S I	90
	orf128ng	L S N I Y L G F R L G Y F D L S A D E N P V L H I W S L A V E E Q Y Y L Y P L L I F C Y K K T K S L R V L R N I S I	172
60	orf128.pep	I L F L I L T A S S F L P S G F Y T D I L N Q P N T Y Y L S T L R F P E L L A G S L L A V Y G O T Q N G R Q T A N G K	150
	orf128ng	I L F L I L T A S S F L P A G F Y T D I L N Q P N T Y Y L S T L R F P E L L V G S L L A V Y G O T Q N G R Q T E N G K	232
65	orf128.pep	R Q L L S L C F G A L C L E V I D K H N P F I P G M T I L L P C L L T A L L I R S M Q Y G T L P T R I L S A S P I	210
	orf128ng	R Q L L S L C F G A L C L E V I D K H N P F I P G M T I L L P C L L T A L L I R S M Q Y G T L P T R I L S A S P I	292

5 The complete length ORF128ng nucleotide sequence <SEQ ID 833> is:

1	ATGCAGACCTG	TGGCATGACG	GCTTGTGAA	GAACGGATTG	GGCCCGGTG
5	CGTCGTCTCC	GTCTATTATT	TCCACCTGAA	TAACCGCTC	CTGGCCGGAGA
10	51	GATTCTGGG	GGTGGACATT	TTCTTGTCA	TCTGGGGATT
101	101	101	101	101	101
15	151	151	151	151	151
20	151	151	151	151	151
25	151	151	151	151	151
30	151	151	151	151	151
35	151	151	151	151	151
40	151	151	151	151	151

This encodes a protein having amino acid sequence <SEQ ID 834>:

45		1 MQAVYRPEI DGLRAVAVIL VIIIFHLNNRW LPGGFLGVDI FFIYISGFLLIT 51 NIILSEQNG SFSFRDYEYT RIRKRYPAET AAVSLASVIO QYIYEDDN 101 QMRKTEILT SFSRNLINYLGF RLGYDFLSDA ENFVLHWSL AVEEYQYLLY 151 PSLLIFCYKT TKSRLVRNIA SIILFLILTA SFSLPAGFYD DLNQPNTYT 201 LSLTRPFELL VGSLLAAVQY TQNQRQTNEN GRKLRSLLS FGALLCVLFV 251 IDKHDFPTG ITFLPLCFTT ALLIRSSYQY TLPIRTRLSAS PFWGKRYIS 301 SLYLYWHIFI AFAYHTIGDK QOLGLPAVASV AALTAGFSL SSYIIEQPLR 351 KRKMFKTFKAF FCYLWASLIM LVGYNLYRSRL IKLGKEHRLP PLPVAAEEN 401 FPFETVLTLCG SHAGHLRGLP DVYGGRECWKR AKLILSDLC SEVWDDEKLAD 451 NFPLCRKYRD VEKAEEAVY QFYDVLRMGGG VPVRPFAEQSF LPGKFARLW 501 ETVKTRIAAV PVYVFANNTS ISRSPLREEK LKRFAINQL RPIRAMGDIG 551 KSNDAVFDLV KDIINPHEWVQ YKQVLPKNTV EIHHGRYLYGD QDHLYFGSY 601 YMGRBTHKHE RLLKHSRGGA I.Q.
50		
55		

ORF128ng and ORF128-1 show 95.7% identity in 622 aa overlap:

60	orf128-1.pep	MQAVRYRPEIDGLRAAVALSVMFHLLNRRWLPGGFLGVDFIFFV1SGLIT1GII1SE1QNG
	orf128ng	MQAVRYRPEIDGLRAAVALSVVIIFHLLNRRWLPGGFLGVDFIFFV1SGLIT1I1SE1QNG
	orf128-1.pep	SFSFRDFTYTRIKR1YPAFIAAVSLASVIASQ1FLYEDFNQMRKTVELSAVFLSN1YLGF
	orf128ng	SFSFRDFTYTRIKR1YPAFIAAVSLASVIASQ1FLYEDFNQMRKTIELTSPVLSN1YLGF

	orf128-1.pep	QQGYFDLSADENPVLIHWSLAVEEQYLLPLLIIFCCKKTKSRLVLRNISIILFLILTA : : : : : : : : : : : : 5 orf128ng RLGYFDLSADENPVLIHWSLAVEEQYLLPLLIIFCCKKTKSRLVLRNISIILFLILTA
	orf128-1.pep	SSFLPSPGFTYTDILNPNTTYYLSTLRFPELLAGSLLAVYGGTQNGRRTANGKRQLSSLC : : : : : : : : : : : orf128ng SSFLPAGFTYTDILNPNTTYYLSTLRFPELLVGSSLAVYGGTQNGRRTENGKRQLSSLC
10	orf128-1.pep	FGALLACLFVIDKHNPFIPGMPTLLEPCILLTALLIRSMQYGTLPTRILSASPIVEVGKISY : : : orf128ng FGALLVCLFVIDKHDFFIPGITLLEPCILLTALLIRSMQYGTLPTRILSASPIVEVGKISY
15	orf128-1.pep	SLYLYIHWIFIARFAHYITGDQKLQLGLPAVSABAALTAGFSLLSYYLIEQPLRKRMTPKKAF : : orf128ng SLYLYIHWIFIARFAHYITGDQKLQLGLPAVSABAALTAGFSLLSYYLIEQPLRKRMTPKKAF
20	orf128-1.pep	FCLYLAPESSLILWGVNLNYARGILLKQEHLRPLFPGAPLAENHEPFTVLTLDGSAGHLRGFL : : orf128ng FCLYLAPESSLMLWGVNLNYSRGLLKQEHLRPLFPGTVAEEENNHFETVLTLDGSAGHLRGFL
25	orf128-1.pep	DYVGSGREGWKAKILSLDSECLVWVDEKLADNPLCRKYRDEVKAEEAVFIAQFYDLRMMGQ : : orf128ng DYVGSGREGWKAKILSLDSECLVWVDEKLADNPLCRKYRDEVKAEEAVFIAQFYDLRMMGQ
30	orf128-1.pep	RPIQAMGDIGKSQNQAVFDLKIIDPVNHWVDAQKYLPKNTVEIYGRYLYGDQDHLYTFGSY : : orf128ng RPIRAMGDIGKSQNQAVFDLKVDPVNHWVDAQKYLPKNTVEIYGRYLYGDQDHLYTFGSY
35	orf128-1.pep	YMGRETHKHERRLLKSHRGALQK : : orf128ng YMGRETHKHERRLLKSHRGALQK 610 620

In addition, ORF218ng shows homology to a hypothetical *H.influenzae* protein:

40	sp P43993 Y392_HAEIN HYPOTHETICAL PROTEIN HI0392 >gi 1074385 pir B64007 hypothetical protein HI0392 - Haemophilus influenzae (strain Rd KW20) >gi 1573364 (U32723) <i>H. influenzae</i> predicted coding region HI0392 [Haemophilus influenzae] Length = 245 Score = 239 bits (604), Expect = 3e-62 Identities = 124/225 (55%), Positives = 152/225 (67%), Gaps = 1/225 (0%)
45	Query: 38 VBFIVISGFLITNIILSEIQNGSFSFRDEYFTRIKRIYFXXXXXXNXXXXXFLYF 97 +DIFIVISGFL I++EIQ STS + FYTRIKRIYF P+Y Sbjct: 1 MDIFFVISGFLITGIITIEQQNSFSLQKFYTRRIKRIYPAFITVMALVSFIASAIFIYN 60
50	Query: 98 DFNQMKYTIELSTVFLSNYLGFRLGYFDLSDAENPVLIHWSLAVEEQXXXXXXKXIC 157 DFN++RKTIEL+ FLSN YLG GFYDLSA+ENPVLIHWSLAVE Q I Sbjct: 61 DFNKLRKTIELIAFLSNFLYGLTQGYFLDSANFVLPVLHWSLAVEQQYLYIFPLILIA 120
55	Query: 158 YKKTKSRLVRNISIILFLILTASSFLPAGFTTDILNPNTYYLSTLRFPELLVGSLLAV 217 YKK + ++V+ I++ILF IL A+SF A FT ++L+QPN YYLS LRFPELLVGSLLA+ Sbjct: 121 YKKFREVKVLFIITLILFILLATFVSANFYKEVLHQPNIYLSNLRFPELLVGSLLAI 180
60	Query: 218 YGQTQNQRRQTEENGKRQLLSCFGALLVCLFVIDKHDFFIPGIT 262 Y N + Q +L+L L CLF+++ + FIPGIT Sbjct: 181 YHNLSEN-KVQLSKQVNNIILAILSTLLFSCLFMNNNIAFIPGIT 224

This analysis, including the identification of several putative transmembrane domains, suggests that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 99

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 835>:

```

5   1 .. ATTAATTACG AACACCGCIG GATGTTTCTT TACGGCCACG TGACGACCTT
  51  GGGCGTGAAC GTCGTGGCAA C.CGGGGGGG TTTCGTTATTC GGTCCTGTTG
101 101 TGGCGTTGGC GGCGCTGATT CACTTGAAA AAGCCGGTGC GCGCGATGGGC
151 151 GTCTGCGGT GGGCGTTGGC TAAAGTTTCG CTGCTGTATG TTACGCTGTT
201 201 CGGGGGTAGC CGCGCTGTTG TGCAAGATTGT GATTTGGGG TATGTGGGTT
251 251 TTCCGTTTTT CGTC.. .

```

This corresponds to the amino acid sequence <SEQ ID 836; ORF129>:

```

10 10 1 .. IYYEYRWML YGALTTLGLT VVAXAGGSVL GLLLALARLI BLEKAGAPMR
  51 51 VLAVALRKVS LLYVTLFRT PLFQIVIWA YWFFFFV..

```

Further work revealed the complete nucleotide sequence <SEQ ID 837>:

```

15 15 1 ATGGATTTC GTTTGACAT TATTACGAA TACCGCTGGA TGTTTCTTA
  51 51 CGGGCGACTG AGCACCTTG GGCTGACCGT CTGCGAACGG GCAGGGGGTT
101 101 CGGTATTGGG TCTGTTGTTT GCGPTGGCGC GCCTGATCA CTGGAAAAAA
151 151 GCGGTGCGC CGATGCGCGT GCTGCGGTGG CGCTTGCGTA AAAGTTTCGCT
201 201 GCTGTAGTTT ACAGCTGTTCC GGCGTAGGCC GCTGTTTGTG CAGATTTGTA
251 251 TTGGCGCTA TGTGTTGTTT CGCTTCTTC TCCATCTTC AGACGGCATT
301 301 TTGGTAGCGO GGGAGGGGG AAATOGCGCT CGTCGGCGAT ACGGGGCGCT
351 351 GATTGCGGT TCTTGGCA TGTCGCCCCA CTGGGGGGG TATATCTGTG
401 401 AGATTTTCCG CGCGGGCATC CAGCTTATAAG ACAAAAGGACA GATGGAGGGCG
451 451 GCGCGTTCTT TGGGGCTGAC CTATCGCGAC CGCGATGGGT ATGTGATTCT
501 501 GCGCGAGCGA TTGGGGCGCA TGCTGCGGCC TTGGGGAGC GAGGTCATCA
551 551 CGCTCTTGA AGAACAGCTG CTGCTGCGC TCATTGCTG TGCGGAGITG
601 601 GCGCTATGTC AGAAATCGAT TAAGGGGGCGA TATTGGGTTT ATGAGAACCC
651 651 GCGTTAACCC GTCCGCGTGA TTATCTGTT GATGACGACT TTCTTAGCT
701 701 GGATATTCTC GGGTTTGGAA AACGTTACA ATCCGCAACA CGCGCTGA

```

This corresponds to the amino acid sequence <SEQ ID 838; ORF129-1>:

```

30 30 1 MDFRFDDIIYE YRWMFLYGA TTLGLTVVAT AGGSVLGLLL ALARLIHLKE
  51 51 AGAPMRVLA WLRKVSLLYV TLEFRGTLFV QIVIWAYWF FFFVHPHSIGI
101 101 LVSGEAIAAL RRGYGPILAG SLALIANSGA YICEIFRAGI QSIDRGQMEA
151 151 ARSLGLTPQ AMRVYIPLCQ LRRMLPFLAS EFTILKLKDSS LLSVIAVVEL
201 201 AYVNQNTITGK YSVYEEFLYT VALIYLMTT FLGWIFIRLE KRYNPGHR*

```

Computer analysis of this amino acid sequence gave the following results:

35 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF129 shows 98.9% identity over a 88aa overlap with an ORF (ORF129a) from strain A of *N.meningitidis*:

40 orf129.pep orf129a	10 IIYEYRWMLYGA 10 MDFRFDDIIYEYRWMLYGA 10 10 20 30 40 50 20 20 30 40 50
45 orf129.pep orf129a	60 ALRKVSLLYVTLFRGTLFV 60 ALRKVSLLYVTLFRGTLFVQIVIWAYWF 70 80 90 100 110 120
50 orf129a	130 SLALIANSGAYICEIFRAGIQOSIDRGOMEARSLGLTYPOAMRYVILPQA 140 150 160 170 180

The complete length ORF129a nucleotide sequence <SEQ ID 839> is:

```

1 1 ATGGATTTC GTTTGACAT TATTACGAA TACCGCTGGA TGTTTCTTA
  51 51 CGGGCGACTG AGCACCTTG GGCTGACCGT CGTCGGCGAT CGGGGGGGTT

```

101 CCGTATTGGG TCTGTGTTG GCCTGGGGC GCCTGATTCA CTTGGAAAAA
 151 GCCGTCGCG CGATGCGGT GCTGGCGTG GCGTGGGTA AGTTTCGCT
 201 GCTGTAGTT ACGCTGTCC GGGGTACGCC GCTGTTGTC CAGATTGTGA
 251 TTGGGCGTA TGTTGTTT CCCTTTTCC PCATCCTTC AGACGGCATT
 301 TTGGTAGCG CGGAGCGGC AATCCCGCTG CCGCGGGAT ACGGCGCT
 351 GATTGCGGT TCTTGGCGCA TGATGCCAA CTCGGGGGG TATATCTGTG
 401 AGATTTCGCG CGCGCGCATG CAGCTCTATAG ACAAAAGACA GAATGGAGCG
 451 GGCGCTTC TTGGGCGTGA CTATCCGAG CGGATGGGT ATGTGATCT
 501 GCGGAGGCA TTGCGCGTA TGCTGCCCGG TTGGCGAGC GAGTTCTATCA
 551 CGCTCTGAA AGAACAGCTG CTGCTGTGCG TCATTCGTTG GGCGGAGTGT
 601 GCGTATGTC AAATAACGAT TACGGGGGG TATTCCGGTT ATGAAAGAAC
 651 GCTTACACC CGTSCCGCTGA TTATCTGTT GATGACGACT TTCTTAGGCT
 701 GGATATTCG CGGTTGGGAA AAACTTACAA ATCCGAACA CGCGCTGA

This encodes a protein having amino acid sequence <SEQ ID 840>:

15 1 MDFRFIDIYE YRWMFLYGAL TTLGLTVVAT AGGSVLGLLL ALARLHILEKAGAPMRVLAW
 51 1 AGAPMRVLAW ALRKVSLLYV TLFRGTLPLFV QIVIMAYVWF PFFVHPHSQDGI
 101 1 LVSGEAAIAL RRGVYPLIG SIALIANSGA EYIEIFRAGI QSIDKGQMEA
 151 1 ARSLGLTPQ AMRVYLIPQG LRMLPPPLAS EFTILKLDSS LLSVIAVAEL
 201 1 AYVQNTITGR YSVYEEPLYT VALIYLMTT FLGWIFLRLK KRYNPQHR*

20 ORF129a and ORF129-1 show 100.0% identity in 248 aa overlap:

orf129a.pep	MDFRFIDIYEYRWMFLYGALTLGLTVVATAGGSVLGLLLALARLHILEKAGAPMRVLAW
orf129-1	MDFRFDIYEYRWMFLYGALTLGLTVVATAGGSVLGLLLALARLHILEKAGAPMRVLAW
orf129a.pep	ALRKVSLLYVTLFRGTLPLFVQIVIMAYVWF PFFVHPHSQDGLIVSGEAAIALRGGYGPILAG
orf129-1	ALRKVSLLYVTLFRGTLPLFVQIVIMAYVWF PFFVHPHSQDGLIVSGEAAIALRGGYGPILAG
orf129a.pep	SIALIANSGAYICEIFRAGIQSIDKGQMEAARSGLLTYPQMRYVILPQALRRMLPPLAS
orf129-1	SIALIANSGAYICEIFRAGIQSIDKGQMEAARSGLLTYPQMRYVILPQALRRMLPPLAS
orf129a.pep	EFTILKLDSSLSSVIAVAELAYVQNTITGRYSVYEEPLYTVALIYLMTTFLGWIFLRLK
orf129-1	EFTILKLDSSLSSVIAVAELAYVQNTITGRYSVYEEPLYTVALIYLMTTFLGWIFLRLK
orf129a.pep	KRYNPQHRX
orf129-1	KRYNPQHRX

40 Homology with a predicted ORF from *N.gonorrhoeae*

ORF129 shows 98.9% identity over a 88 aa overlap with a predicted ORF (ORF129ng) from *N.gonorrhoeae*:

orf129.ng	IIEYEYRWMFLYGALTLGLTVVAXAGGSVLGLLLALARLHILEKAGAPMRVLAW	54
orf129ng	MDFRFDIYEYRWMFLYGALTLGLTVVATAGGSVLGLLLALARLHILEKAGAPMRVLAW	60
orf129.ng	ALRKVSLLYVTLFRGTLPLFVQIVIMAYVWF PFFVILHTAF	88
orf129ng	ALRKVSLLYVTLFRGTLPLFVQIVIMAYVWF PFFVILHTAF	120

An ORF129ng nucleotide sequence <SEQ ID 841> was predicted to encode a protein having amino acid sequence <SEQ ID 842>:

55 1 MDFRFIDIYE YRWMFLYGAL TTLGLTVVAT AGGSVLGLLL ALARLHILEKAGAPMRVLAW
 51 1 AGAPMRVLAW ALRKVSLLYV TLFRGTLPLFV QIVIMAYVWF PFFVILHTAF
 101 1 LGNAMRQSRV VFDKGWIAKG SLELNQCPRG RKTGRFPPG ESNLGTEPRN
 151 1 PLSMQRFF GCEMWYFPQN KF*

Further work revealed the following gonococcal sequence <SEQ ID 843>:

1 ATGGATTTc gTTTGACAT TATTTAcgaA TACCGCTGGA TGTTCCTTA

51 CGCGGCACTG ACGaccgttg ggcttacggct ctggggcgac gggGGCCGAAATT
 101 CGCTatggc tg CTGTGTTGTC GCGTTGGCGC GCTGGATTCATA CTGGGGAAATT
 151 GCGCGTGGCGC GAATGGCGC GCGTGGCTGG GCGTGGCTGTA AGATTTTGCAG
 201 GCTGTAAGCTT ACCCTGTTGTC GGGGGTAGCCCG GCTGTGGTTCAG CAGTTTGCTGAG
 251 TTGGGCGCTGA TGTTGTTGGT CGCTTGGTTTCG TCCATCTTCCTG AGACCGCATG
 301 TTGGTCAGCGC GAGGGCGGCC AAATCCGGCTG CCGGGGGATG AGCGGGCGCTG
 351 GATTGCTTCCGCTT TTGGTGGCG TGATGGCCAA CTGGGGGGCTA TATAGCTGGC
 401 AGATTTTCCGG CGCGCCGACAT CAGCTTATAG AACAAAAGGACA GTATGGCTGCG
 451 GCGTGTCTT GGAGCTGAC CTATCCGGCG ACCTGGGGCTG ATGATGATTCAT
 501 GCGCGGAGCGA TTGGCGCTGA TGCCTGGCGC TGGTGGGGAGT GAGTTCTCATG
 551 CGCTGTCTGA AGAACAGCTC CTGCTGTCG TGATCTGGCTG GGGGGAGGTT
 601 GCGTATGTCG AGAATACGAT TACGGGGCGG TATAGCTGGT ATGAGAACGCT
 651 GCTTAACTACC CGGSCCCCTG TTTATCTGGT GATGSGAGCT GCTTCTAGGCT
 701 GGATATTCCTT CGGTGTTGGA AACACCTTACA CCTGGACCAA CGCTGGCT

- 15 This corresponds to the amino acid sequence <SEQ ID 844; ORF129ng-1>;

```

1 MDFRDLEIYR YRWNPFLYGLA TTLGLTIVVAT AGGSVLIGLL ALARLHILEK
51 AGAPFRVLAW ALRKVSLILV RRFGRTPFLV QIVIWAYVWF PFFVPHSDGIE
101 LVSGEAAIAL RGGYGPGLIA SLALIANSGA YICEIF FRAGI QSIIDKGOMER
151 ARSLSLLTYPQ AMVRVLPQ LRRMLKPLA EFTILWDKSS LLSVIAVAEEL
201 AVQVNNTIGR YSVVEEPYLTT VALIYLMLLT FLGWFLRLRE KYNQRPH

```

ORF129ng-1 and ORF129-1 show 99.2% identity in 248 aa overlap.

	orf129-1.pep	MDFRFDIYYEWMPFLGALTTIGLTVVATAGGSVLGLLLALARLIHLEKAGAPMVRVLAW
25	orf129ng-1	MDFRFDIYYEWMPFLGALTTIGLTVVATAGGSVLGLLLALARLIHLEKAGAPMVRVLAW
	orf129-1.pep	ALKRVSLLYVTLFRGTPLFQIVIWAYWFPFFVHPSDGILVSGEAAIAIRRGYGPILAG
	orf129ng-1	ALKRVSLLYVTLFRGTPLFQIVIWAYWFPFFVHPSDGILVSGEAAIAIRRGYGPILAG
30	orf129-1.pep	SLALIANSGAYICEIFRAGIQSIDKGQMEEAARSLLGTYPQAMRYVILPQLARLMLPPLAS
	orf129ng-1	SLALIANSGAYICEIFRAGIQSIDKGQMEEAACSLLGTYPQAMRYVILPQLARLMLPPLAS
35	orf129-1.pep	EFTTLLKDSSLSSLVIAV рЕLA YV QNTITGRYSVYEELPYTVALIYLMMTFLGWIFIRIE
	orf129ng-1	EFTTLLKDSSLSSLVIAV рЕLA YV QNTITGRYSVYEELPYTVALIYLMMTFLGWIFIRIE
	orf129-1.pep	KRYNPQHXR
40	orf129ng-1	KRYNPQHXR

In addition, ORF129ng-1 is homologous to an ABC transporter from *A. fulgidus*.

2650409(AE001090) glutamine ABC transporter, permease protein (glnP)
 [Archaeoglobus fulgidus]Length = 224
 Score = -132 bits (329), Expect = 2e-30
 Identities = 86/178 (48%), Positives = 103/178 (57%), Gaps = 18/178 (10%)
 45 Query: 65 VSLLYVTLFRGPTPLFVQIVINAYWVWFPPVHPSDGLILVSGEAAIAIRRGYGPPLAGSLAL 124
 +S YV + RGTPL VQI+-----+ F P+ GI + E A G +AL
 Sbjct: 58 ISTAYEVIRGTPPLVQIIL-----VYFGGLPAIGINLQPFEPA-----GIAL 99
 50 Query: 125 IANSGYACEIFRAGIQSIIDKGQMEEACSLTGQYQAMRYVILPQLARRMLPPIASEFIT 184
 SGAYI ET RAGI+SI GMEAE SLG+TY QAMRYVII PQA R +IP L +EPI
 Sbjct: 104 SICSGYACEIIRVAGIESIPIGMEAAARSLGTMYQYAMRYVIFVQPAFRNLIPLGNGIEI 159
 55 Query: 185 LLKDSLSLSSVIAVAELAVQNITGRYSVVEEPYLTAALIYLLMTFLGLWIFLRLERK 242
 LLKDSLSLSSV+ EL V I + + + AL YL +MT L + +K+
 Sbjct: 160 LLKDSLSLSSVTSVTEVTRGVQYNTTENATPEFLGVALYVLMETLPSRIVASVOKS 217

This analysis, including the identification of transmembrane domains in the two proteins, suggests that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 100

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 845>:

```

1 ..CTGAAAGGAT GCGCTCTGAA AGACCCCTGT TTTATTCCAA ATATCCTTAA
5 TAAGAACATC GCGCATTACTT TCCTGCTCTA CGACCGGGCG GCGGAACATT
10 GGCTGCCGC CGAAACCGCC GGTTTACCC CGCTCGCGGT CGGCCTTCATC
15 CTGCTGCCGC AGCTGCGTG gCTTCAACCA CACCAACTCT TAGCTAAACA
20 cTACCTCGGC ACTTAATCay TGCTCCAACT CTTTGC CGGC GCAGggTaGt
25 TTGTGGAG GCGGGCGWA ATTACAAAAC CTTGGCGGCY CGCGGCCCT
301 GCACCTGATT ACCCTCGGC CGATGATGG CGCGCTGATG ATGGTGTGGC
351 TGACCCCGGC AGCTGCGAC ACCCGCTTAA CCRAAATCGA CTACCCAAA
401 CTCTGCCGA TTGGCGGCC CATCTTTC CGCCGCCCGG TCTCGCGGC
451 TTCTCTGTTG AACGGAGACC CGTATTTT CATTACCGTT CCTGGGATTC
501 TGACCCCGGC CSTATGCGTA CTGTATCTT TCGCTGTTAT ACCGATATT
551 CGGGCGAATG CTTTTACAGA CGATCGGA TAR

```

15 This corresponds to the amino acid sequence <SEQ ID 846; ORF130>:

```

1 ..LKECRKLKDPP FIPNIVVKNI AITFLLLHAA AELMLPAQTA GFTALAVGFI
5 LLAKLRLHHH HELLRKHVYR TYYLQLFEE AGSLWTGAAK LQNLPAASPL
10 HLTILCGNMG GVMMVWLTAG LWHSGFTKLID YPKLCLRIAVP ILFAAAVSRA
151 FLXVNPNPFX ITVPAILTAA FVFLYLFXFI PIFRANAFDT DPE*

```

20 Further work revealed the complete nucleotide sequence <SEQ ID 847>:

```

1 ATGCGGGCGT TTTTCGTGCG CGGGCGGTG CTGGCCATAC TCGGTGCGCT
51 GTGGTGTTC ATCACCCCCG GTGCGCATGTT CCTGCACCG CAATAATTCT
101 TTGGAACTTAT GCTGCCGGCG CGATCGGGC GTTTTTGAC TCGGCCTTGT
151 TTGGACTGGA CGGGTTTTTC GGTTAACTC AAACCTGTCG CGACTTTGAT
201 GGGCGCAITA TTGGCTCGCG CATCGCTAT ACTGCCCCTT TCGCCGCCAA
251 CTGGCTCTTT TTTCGTGCG GCTTATTGGC TTGGTGTGTCG GTCTGTTCTGC
301 GCGCCGGCTGA TTGGCTTAGA CGGAAACACC GACAACCTTG CCTCTGTAAT
351 GTTACTATGCC GGGTCACTCG TTTTCAGACG GCGATATGCC GTACGGGGC
401 ATTGAACTCCG GTGGCGGGC CAAGTGCATC TAAATATGGC GGCGGGTGTG
451 TTGGTATCCG TGCGCGTCAG TATTTTTTCG GCGCGGAAG CCTGAAAGA
501 ATGCGCTGC AAAGACCTCG TTTCATTTCC AAATATCGT TATAAAAACAA
551 TCCCATATAC TTTCCTGCTC TTGGACGCC CGCGGAACG TTGGCTGCC
601 GCGCAACACG CGGGTTTAC CGCGCTGCCG GTCGCGCTCA TCTCTGCTGC
651 CAAGCTGGT GAGCTTCAC ATCACGAACT TTACGTAAA CACTAGTCC
701 GCACTTAITA CCTGCTCGCA CTCTGGCG CGCAGGCTA TTGGTGGACA
751 GGGGGGGGA AATACAAGAA CCTGCGGCC CGCGGCCCGG TGCACCTGAT
801 TACCTCTGGC GCGCATGTTG CGCGCGTGAT GATGGTGGT CTACGGCG
851 GACTGTGGCA CAGGGCGTTT ACCAAACCTCG ACTACCCCAA ACTCTGCGC
901 ATGGCGTFC CCAACCTTT CGCGCGGCC GTCTCGGGG TTTCCTGAT
951 GAACTGGACG CGGTATTTT TCATCACCGT TCTCGGGATT CTACCGCCG
1001 CGGTATGCTG ACTGTATCTT TCACGTTTA TACCGTAT TCGGGCGAT
1051 GCGTTACAG ACGATCGGA ATA

```

This corresponds to the amino acid sequence <SEQ ID 848; ORF130-1>:

```

1 MRPFFVGRAV LTALGALVFF INPGAVLHR QIFLELMPLA AYGGFLTAAL
51 LDWTFGSQNL KVFTIMAAAL LLAAASALPF SFGTASEFVA AYNLVLLFC
101 ARLIWLDRNT DNFAFLMLLA AFTVFQFAYA VSGDINLLRA QVHLNMRARVM
151 FFSVRVSIIL GEAALKERCL KDFVFPINIV YKNTIAITFL LHAAAEELNLP
201 AQTAGFTALG VGFILLAKLR ELHHHELLRK HYVRTYLLQ DFAAGAGLYWT
251 GAAKQNLPAF SAPLHLITLG GMGGGVMMVVW LTAGLWHSGF TKLDYFKLRC
301 IAVEFLFAA VSRAFLMNNV PIFFITVPAI LTAAVFVLYL ETFPIPIFRAN
351 AFTDDE*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF130 shows 94.3% identity over a 193aa overlap with an ORF (ORF130a) from strain A of *N.*

55 *meningitidis*:

-463-

		10	20	30
	orf130.pep	IKECRKLDPFVIPNIVYKNNIAITFLLLHAA		
5	orf130a	LNLRLAQVHLNMAAVMFVSVRVSILLGAELKECRKLDPFVIPNIVYKNNIAITFLLLHAA		
		140 150 160 170 180 190		
	orf130.pep	40 50 60 70 80 90		
10	orf130a	AELNLPAQTAGFTSLAVGFIILLAKLIRELHHHELLRKHYRTYYLLQLFAAAGSLWTGAAX		
		200 210 220 230 240 250		
	orf130.pep	100 110 120 130 140 150		
15	orf130a	LQNLPASAPLHLIYLGGMMGGSVMWLTAGLNHSGLPKLCRIAVPILFAAAVSRRA		
		260 270 280 290 300 310		
	orf130.pep	160 170 180 190		
20	orf130a	FLXNVNPXFFITVPAILTAAVFVLVLFXXFIPFRANAFDDEPX		
		320 330 340 350		

The complete length ORF130a nucleotide sequence <SEQ ID 849> is:

25	1	ATGCCGCCGT TTTGGCGTCC
	51	GGTGTTTTC ATCACCCCCG GTGCCATCGT CTCGCACCGC CAAATTTCT
	101	TGGACTATAT GTCGCCGGGG GCATACCGCG GCTTGTTTGAC TGCGGTTTG
	151	TTGGACTTGG CGGGTTTTC GGTTAACCTTG AAACCGCTTG CGACITTGAT
	201	GGCCGCAATTA TTGGCTGCCG CATCCCCATG ACTGOCCTTG TCCCGGAAA
	251	CTGCCCTGTT TTGCTGCCG CCTATTGGC TGGTTGCT GCTGTTCTGC
	301	GCCCCCGTGA TTGGCTAGA CGCATATTGGC GACACTCTG CCTGCTATA
	351	TTGACTTGG CGGTCTCACG TTTCAGACAG GUCATATGGC GTACGGGGCG
	401	ATTGAACTT GTTGGGGGG CAAGTCATC TAAATATGGC GGGGGTGTG
	451	TTGCTATTCG CGGCCCGTCA TATTCTTTCG GGCGCCGGANG CCTCTGAAAGA
	501	ATCGCCCTGAA AAGAACCCCA TATTCTACCG CAAATGTCGTC TATAAAAAAC
	551	TCCGCACTTAC CTTCCTGCTC CTGCAAGCGG CGCGGCAACT TTGGCTGCCT
	601	GCGCAACACC CGGGTTTACG CTGCTTGGCG TCGGGCTTA TCCCTGTTGC
	651	CAAGCTGGT GAGCTTCACCC ATACGAACTT CTTCGGAAA CACZAGTCC
	701	GCACTTATAA CCTCTCCCA CCTTGTGGCG CGCAGGCTA TTGGTGACA
	751	GGGGGGGAA ATTAAACAAA CCCTGGGGGG TCCCGGGGGCC TTGACCTGAT
	801	TACCTCTGGT GCGCATGATG CGACGCTGAT GATGCTGTGG CTGACTGGCG
	851	GACTGTGGCA CGAGGGCTTT ACCAACGCTCG ACTACCGGAA ACTCTGCGC
	901	ATGCGCGTCC CCACTCTNTT CGCCGGGGG GTTTCGGGGG CTGTTTTAAT
	951	GAACCTAAAC CGGATATCT TCATCACCGT CCCGGAAATT CTGACCGGGG
45	1001	CCGTTGTCGT GCTTACTCG CTGACATTCC TACCGATCTT TCGGGCGAAC
	1051	GCGTTTACAG ACGATCCCGGA ATA

This encodes a protein having amino acid sequence <SEQ ID 850>:

50	1	MRRPFVGAAV LAILGALVFF INPGAIVLHR QIFLEMLPA AYGGFLTAAL
	51	LDWTFGSGNL KPVATIMAA LILAASAILPF SPQATSFVVA AYVLLVLLFC
	101	ARLWLRDNY DNFALIMLA AFTVFTOTAYA VSGDLNLLRA QVHLNMAAVM
	151	FVSVRGAVALL AFTVFTOFVVA YKNTAITFL LHAASEEINLP
	201	AQTAGFTSLA VGFTLAKLRL ELHHHELLRK HYVRTYYLLQ LFAAAGYLWT
	251	GAALKLNLPA SAPLHLITLG GMGGSVMMVW LTAGLWHSGF TKLDYFKLCR
	301	IAPVPLFAA SRSRVALNNN PIFFITVEAI LTAAVFVLV LTFVPIFRAN
	351	AFTDDPE*

ORF130a and ORF130-1 show 98.3% identity in 357 aa overlap:

60	orf130a.pep	MRPFFVGAAV LAILGALVFF INPGAIVLHR QIFLEMLPA AYGGFLTAALLDWTFGSGNL
	orf130-1	MRPFFVGAAV LAILGALVFF INPGAIVLHR QIFLEMLPA AYGGFLTAALLDWTFGSGNL
	orf130a.pep	KPVATLMAALLLAASAILPFSPQATSFVVA YVLLVLLFCARLWLDRNTDNFALLMLA
	orf130-1	KPVATLMAALLLAASAILPFSPQATSFVVA YVLLVLLFCARLWLDRNTDNFALLMLA
65	orf130a.pep	ATPVFTQFAYAVSGDLNLLRAQVHLNMAAVMFVSVRVS ILLGAELKECRKLDPFVIPNV

	orf130-1	: : : : AFTVFQTYAVSGLDNLLRAQVHLNMAVMFVSVRVSILLGAEALKECRKLDPFVIPNIV
5	orf130a.pep	YKNIAITFLLHAAAEIWLPAQTAGFTSLAVGFIALLAKRELHHHELLRKHYVRTYYLLQ
	orf130-1	YKNIAITFLLHAAAEIWLPAQTAGFTSLAVGFIALLAKRELHHHELLRKHYVRTYYLLQ
10	orf130a.pep	LFAAAAGYLWTGAALKLQNLPASAPLHLITLGGMMGSVMMWLTAGLWHSGETKLDYPKLCR
	orf130-1	LFAAAAGYLWTGAALKLQNLPASAPLHLITLGGMMGVMMWLTAGLWHSGETKLDYPKLCR
15	orf130a.pep	IAVPILFAAAVSRAVLMNVNPIFFITVPAILTAAVFLVLYLTFIPIFRANAFDTDPE : : orf130-1
	orf130-1	IAVPILFAAAVSRAFLMNVNPIFFITVPAILTAAVFLVLYLTFIPIFRANAFDTDPE : :

Homology with a predicted ORF from *N.gonorrhoeae*

ORF130 shows 91.7% identity over a 193 aa overlap with a predicted ORF (ORF130ng) from *N.gonorrhoeae*:

20	orf130.pep	LKECRKLDPFVIPNIVYKNIAITFLLHAA 30 : : orf130ng
	orf130ng	LNLLRAQVHLNMAVMFVSVRVSVLLGTETLKECRKLDPFVIPNIVYKNIAIT-LHAA 201
25	orf130.pep	AELWLPAQTAGFTALAVGFIALLAKRELHHHELLRKHYVRTYYLLQLFAAAGSLWTGAAX 90 : : orf130ng
	orf130ng	AELWLPAQTAGFTALAVGFIALLAKRELHHHELLRKHYVRTYYLLQLFAAAGYLWTGAAX 261
30	orf130.pep	LQNLPASAPLHLITLGGMMGVMMWLTAGLWHSGETKLDYPKLCRIAVPLFIPASVRA 150 : : orf130ng
	orf130ng	LQNLPASAPLHLITLGGMMGVMMWLTAGLWHSGETKLDYPKLCRIAVSILFASAVRA 321
	orf130.pep	FLXNVNPFITVPAILTAAVFLVLYLTFIPIFRANAFDTDPE 193 : : orf130ng
	orf130ng	VLMVNPNPIFFITVPAILTAAVFLVLYLTFIPIFRANAFDTDPE 364

An ORF130ng nucleotide sequence <SEQ ID 851> was predicted to encode a protein having amino

35 acid sequence <SEQ ID 852>:

1	MNKFETHMPF PFFVGRAVLA ILGALVFFHO PRRYHPAPPN FLGTYAAGCI
51	RRFEDYRWFV PGDFGDRQPEP CRYFDGGVVA CGCGFIAVET ATCRFRRL
101	LAGAVAVLRL ADLARQHQRT LRSVDVTAAT TVFQTAYAVS GDINLLRAQV
151	HLNLMAMVNFV SVRVSVLLGT ETIIEKCRLK DPFIPWVIYK NIAITLLLHA
201	AAEWLWLPAQT AGTFLALAVGF ILLAKRELH HHELLRKHYV RTYLLQLFA
251	AAGYLWNLPAQT KLNQLNFGASAF LHLITLGQMT GGVMWMLTA GLWHSGFTKL
301	DYPKLCRIAV SILEPSAVSR AVLMVNPNFIT FITVPEILTA AVFMLYLTTF
351	VPIFRANAFDTDPE*

Further work revealed the following gonococcal DNA sequence <SEQ ID 853>:

45	1 ATGCCGCCGT TTTCTGCGG TCGGGCAGTA CTTGCCATAC TCGGTGCGTT
	51 GGTGTTTTT ATTACCAAGCC GGGCTATCACI CCTGCACCGC CAAATTTCCT
101	TGGAACTTAT GCTGCCGGTG GCATACCGCG TTTTTTGAC TACCGCTTTG
151	TTGGACCCGA CGGGTTTTTC AGGCCRACCTG AAACCTGCGG CTACTTTGAT
201	GGCGCTGTG TTCTCTGTC CGCGCTGTTT ATTGCCTTT TAACCGAAC
251	TTGCCCATTT TTCTGCTGCG GCCTATGCGC TGGTGTGTTGCT GCTGTTCTGC
301	GCTGGCTGTA TTTGGCTCGA CGCGCACACC GACAACCTCG CTCTGTTGAT
351	GTTACTTGCCT GCATTTACCG TTTTCAGAC CGCCATTCG GTCAGCGCGC
401	ATTGAGACTT ACTGGCGCGG CAAGTGCATT TGAATATGGC GGCGCTCATG
451	TTGTGATTCG TCCCGCTCGA CGTCTCTTTC CGCACGGAAA CCTGAAAGA
501	ATGCCGCCGT AAAGACCCCG TATTCATCCCA CACCGTTTAC TATAAAAACCA
551	TGCCCATCAC CCTGCTGCTG CACGCCCGC CGAACCTTGC GCTGCCGGCG
601	CAAACGCCCG GTTTTACTGC GCTTGCCTGCG TCCTCTCATCC TGCTGCGCAA
651	GCTGGCGCAA CTGCCCCATC AGCAACTCTT ACGCCAACAC TACGTGCGCA
701	CTTATTAACCT GCTCCAGCTC TTGGCGCCCG CAGGTTATCTT GTGACAGGC
751	CGCGGCCAACAT TGCACAACTT CGCCGCCCTGC ACCCTGATTAC
801	CCTCGCCGGC ATGACGGGTG CGCTGATGAT GGTGTTGGCTG ACTGCCGGAC
851	TGTGGCACAG CGGCTTTTAC AACATCGACT ACCCGAAACT CTGCGGCATC

901 GCGGTCTCCA TCCCTTCGCG CTCGGCGCGTT TCGCGGGCTG TTTTAATGAA
 951 CGTGAATCGG ATATTCCTCA TCACCGTTCG CGAGATTCTG ACCGCGCCG
 1001 TGTTCACTGCT TTAACCTGCTG ACGTTCTGAC CGATTTTG AGCGAACCG
 1051 TTTCACAGACG ATCCGGATA A

- 5 This corresponds to the amino acid sequence <SEQ ID 854; ORF130ng-1>:

1 MRPFFVGAAV LAILGALVFF INPGAIILH P QIFLELMPLA AYGGFLTTAL
 51 LDRTFGSGNL KPAATLMAVL LIVAAVLLPF LPQLAAPFVA AWYLWLLLFC
 101 AWLILDRNT DNFALLMLLA AFTVFOATAVA VSGDLNLLRA QVHLNMAAVM
 151 FVSVSVSLL GTETLKECR L KDPFVIPNVI YKNAITATLL HAAAEELMLPA
 201 CTAGTATLAV GFILLAKLRE LHHHEFLRKH VYRTVYLLQL FAAAAGYLWTG
 251 AAKLQNLPAS APLHLITLG NGTGVNMWV TAGLWHSGET KLDYPKLCRI
 301 AVSILFASAV SRAVLNVNP IIFTITVPEIL TAAVTMLYLL TFVPIFRANA
 351 FTIDPE*

ORF130ng-1 and ORF130-1 show 92.4% identity in 357 aa overlap:

15	orf130-1.pep	MRPFFVGAAVLAILGALVFFINPGAIVLHRQIFLEIMLPAAAYGGFLTAALLDTGFSGNL
	orf130ng-1	MRPFFVGAAVLAILGALVFFINPGAIILH RQIFLEIMLPAAAYGGFLTTALLDRTGFSGNL
20	orf130-1.pep	KPVATIMALLAASAII LFPSPQTAFVVAAYWLVLLFCARLIWLDRNTDNFALLMLLA
	orf130ng-1	KPAATIMAVLVAVALLPFLPQLAFFVAAYWLVLLFCAWLIWLDRNTDNFALLMLLA
25	orf130-1.pep	AFTVFQTAYAVS GDNLNLLRAQVHNNAAVMFVSVRVSIILGAELKECRKLKDPTVFPNIV
	orf130ng-1	AFTVFQTAYAVS GDNLNLLRAQVHNNAAVMFVSVRVSVL LGTETLKECRKLKDPTVFPNIV
30	orf130-1.pep	YKNAITFLLHAAAELWL PAQTAGFTALAVGFILLAKLRFLLHHHEFLRKH VYRTVYLLQ
	orf130ng-1	YKNAIT -L LHAAAELWL PAQTAGFTALAVGFILLAKLRFLLHHHEFLRKH VYRTVYLLQ
35	orf130-1.pep	LFAAGGYLWTGAALKLQNLPSAPSLHLITL LGGMGMMGVWL TAGLWHSGFTKLDYPKLCR
	orf130ng-1	LFAAGGYLWTGAALKLQNLPSAPSLHLITL LGGMGMMGVWL TAGLWHSGFTKLDYPKLCR

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

40 Example 101

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 855>:

1 ATGGAAATTC GGGCAATAAA ATATACGCCA ATGGCTCGCT TGCTTGATT
 51 TACGGTTGCA GGCTGGCGGC TGCGGGGGT GTATGAGTC TGCTCCCTCA
 101 CGCGTTGGTG TAAGCGCAGA AACCGCGCTC CCATGAGATT TTGGGATATT
 151 GGCGGGAGAG GTCGGCGCTC TTAAAGGGAC CAGAGATAC CGCTTCCAGA
 201 CGCCAAATGTG TCCCTCACGG CAACCGAATA TGAATCCGCA CAAACATCTT
 251 ACTTTACAG GAAATTAAGG AACATTGAGG C.TGGGGCT GTATGGCT
 301 AAGCGTGAAG GCACACCTTT GATTGAGACG TTCAACAGG GAGGATTG
 351 CTGCTTGGAA AAG .

- 50 This corresponds to the amino acid sequence <SEQ ID 856; ORF131>:

1 MEIRAIKYTA MAALLAFTVA GCRLAGWYEC SSLTGWCKPR KPAIADFWDI
 51 GGESPSSPLGD YEIPLSDGS SVRANEYESA QCSFYRKIG KFEXXGLWR
 101 TRDGKPLIET FKQQGGFDCLKE ..

Further work revealed the complete nucleotide sequence <SEQ ID 857>:

55 1 ATGGAAATTC GGGCAATAAA ATATACGCCA ATGGCTCGCT TGCTTGATT
 51 TACGGTTGCA GGCTGGCGGC TGCGGGGGT GTATGAGTC TGCTCCCTCA

101 CGCGCTGGT AAAGCCGAGA AACGGCGCTG CCATCGATT TGGGATATT
 151 GGCGGCGAGA GTCCGGCGTC TTAAAGGGAC TAGCAGGATC CGTTTCTCAGG
 201 CGCAATCGT TCCGTCAGGG CAAACGAATA TGAATCCGCA CAAACATCTT
 251 ACTTTACAG GAAAATAGGG AAATTTGAGG CTTCGCGGT GAGTTGGCC
 301 AGCGCTGAG GCAAAACCTT GATTGAGACG TTTAAACAGG GAGGATTGTT
 351 CTGCTTGAA AAGCCAGGGT TCGGGCGCA CGCTTGCTC GAGGCCCTTC
 401 GATGTTAA

This corresponds to the amino acid sequence <SEQ ID 858; ORF131-1>:

1 MEIRAIKYTA MAALLAFTVA GCLRAGWYEC SSLTGWCCKPR KPAADFWDI
10 51 GGSEPPSIGD YEIPLSDGNR SVRANEYESA QCSQFWYRKIG KFEACGLDWR
101 TRDGKPLIET FKQGGFCDLCE KCGGLRRNGLS ERVRW*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF131 shows 95.0% identity over a 121aa overlap with an ORF (ORF131a) from strain A of *N. meningitidis*:

	orf131.pep	MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLTGWCQPKPAIDFWWDIGGESPPSLED
20	orf131a	MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLTGWCQPKPAIDFWWDIGGESPPSLED
		70 80 90 100 110 120
	orf131.pep	YEIPLSDGNSSVRANEYESAQCSFYRKIGKFEACGLDWTRDGKPLIETFKQGGFDCL
25	orf131a	YEIPLSDGNSSVRANEYESAQCSFYRKIGKFEACGLDWTRDGKPLIETFKQGGFDCL 70 80 90 100 110 120
	orf131.pep	K
30	orf131a	KQQGLRRNGLSEWRVRWX 130

The complete length ORF131a nucleotide sequence <SEO ID 859> is:

35	1 ATGGAAATTC GGGAAATAAC 51 TACGGTGTCA CGCTGGCGGT 101 CCGCTGGTG TAAGCCGAGA 151 GGCGCCGAGA GTCTCCCTGC 201 CGGCATCTGC TCCGTCAAGG 251 ACTTTTACAC GAAAATGGAG 301 ACGCTGAC GCAAATTCCTT 351 TTGTTTGAAGA AAGCAGGGGT 401 GATGGTAA
40	

This encodes a protein having amino acid sequence <SEQ ID 860>:

45 1 MEIRAIKYTA MAALLAFTVA GCRLAGWEC SSLSGWCKER KPAADFWDI
51 51 GGESPSSLED YEIPLSIDNGN SVRANEYESA QCSYFTRKIG KFEACGLDWR
101 101 TRDGKPLIET FKQEGFLCKL KQGLRRNGLS ERVRW*

ORF131a and ORF131-1 show 97.0% identity in 135 aa overlap:

	orf131a.pep	MEIRAIKYTAMAALLAFTVAGCRLAGWYCSSLGSWCKPRKPAAIDFWDINGESEPPSLEDE
50	orf131-1	
	orf131a.pep	YEIPLSDQNRNSVRANYEESAQQSYFYRKIGKEFACGLLDWTRDGKELLIETFKQEGFCDCLK
	orf131-1	
55	orf131a.pep	YEIPLSDQNRNSVRANYEESAQQSYFYRKIGKEFACGLLDWTRDGKPLIETFKQGGPFCDC
	KQGLRRNGLSEVRWXX	

orf131-1 KQGLRRNGLSERVRWX

Homology with a predicted ORF from *N.gonorrhoeae*

ORF131 shows 89.3% identity over 121 aa overlap with a predicted ORF (ORF131ng) from

5 *N.gonorrhoeae*:

orf131.pep	MEIRVIKYTAMAALLAFTVAGCRLAGWYECSSLTGWCCKPRKPAIDFWDIGGESPPSLGD	60
orf131ng	MEIRVIKYTATAALFAFTVAGCRLAGWYECSSLTGWCCKPRKPAIDFWDIGGESPLSLED	60
10 orf131.pep	YEIPLSDGNSSVRANEYESAQSQSYFYRKIGKFEXGGLDWRTRDGKPLIETFKQGGFDCL	120
orf131ng	YEIPLSDGNRSVRANEYESAQSQSYFYRKIGKFEXGGLDWRTRDGKPLIETFKQGGFDCL	120
15 orf131.pep	K 121	
orf131ng	KQGLRRNGLSERVRW 134	

A complete length ORF131ng nucleotide sequence <SEQ ID 861> was predicted to encode a protein having amino acid sequence <SEQ ID 862>:

1	MEIRVIKYTA TAALFAFTVA GCRLAGWYEC LSLSGWCCKPR KPAATDFWDI
51	GGESPLSLED YEIPLSDGNR SVRANEYESA QKSYYFRKIG KFEACGLDW
101	TRDGKPLVER FKQEGFDCL E KQGLRRNGLS ERVRW*

Further work revealed the following gonococcal DNA sequence <SEQ ID 863>:

1	ATGGAAATTG GGGTAATAAA ATATACGGCA ACGGCTGCCT TGTTTGCATT
51	TACGGGTGCA GGCTGCCGGC TGCGGGGGTG GTATGACTGT TGTCCTTGT
101	CCGGCTGGTG TAAGCCGAGA AAACCTGGCC CCATGAGATT TTGGGATATT
151	GGGGGGAGAGA GtcccgctGTC TTTAGGAGAC TAGAGATAC CGCTTTCAGA
201	CGGGATCTGT TCGCTCAGGG CAAACGATA TGAATCCGGG CAAAATCTT
251	ACTTTTATAG GAAATAGGG AAGTTTGAAG CCTGGGGGTT GGATTGGCT
301	ACGGTGTGAGC GCAAACCTTT GGTTGAGAGG TTCAACACAGG AAGGTTTCCA
351	CTGTTGGAA AAGCGGGGT TGCGGGCA CGGCGTGTCC GAGCGGCGTCC
401	GATGGTAA

This corresponds to the amino acid sequence <SEQ ID 864; ORF131ng-1>:

1	MEIRVIKYTA TAALFAFTVA GCRLAGWYEC SSLSGWCCKPR KPAATDFWDI
51	GGESPLSLED YEIPLSDGNR SVRANEYESA QKSYYFRKIG KFEACGLDW
101	TRDGKPLVER FKQEGFDCL E KQGLRRNGLS ERVRW*

ORF131ng-1 and ORF131-1 show 92.6% identity in 135 aa overlap:

orf131ng-1.pep	MEIRVIKYTATAALFAFTVAGCRLAGWYECSSLTGWCCKPRKPAIDFWDIGGESPPSLGD
orf131-1	MEIRVIKYTAMAALLAFTVAGCRLAGWYECSSLTGWCCKPRKPAIDFWDIGGESPPSLGD
40 orf131ng-1.pep	YEIPLSDGNSSVRANEYESAQSQSYFYRKIGKFEXGGLDWRTRDGKPLIETFKQGGFDCL
orf131-1	YEIPLSDGNRSVRANEYESAQSQSYFYRKIGKFEXGGLDWRTRDGKPLIETFKQGGFDCL
45 orf131ng-1.pep	KQGLRRNGLSERVRWX
orf131-1	KQGLRRNGLSERVRWX

Based on the presence of a predicted prokaryotic membrane lipoprotein lipid attachment site, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 102

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 865>

```

1   ATGAAACACCA TCCATATTAT CGGTATCGGG GGCACGTTA TGGGCCGGCT
51  TGGCCGCATT GCAGAACGAG CGGGGTGTC AGTCAGCGGT TGGCACGGGA
5101 AGATGATTCGC CGCCGATGAGC ACCCAAGCTCG AAGCCCTTGGG TATAGACGTG
151 TATGAAAGCT TGATGCGGC TCAGTGTGGAC GAATTTAAAG CGCACCTTTA
201 CGTTATCGGC ATATGCCCCA AGCGCGGAA CGGATGTTGGT GAAGCGGATTI
251 TGACCTCGG CCTGCCTAT ATTICCGGCC CGCAATGSGCT GTCGGAAAC
301 GTGCTGCAAC ATCATGGGT ACTCGTGTTG GCGGGGRCG ACGGCAAAAC
351 GACCCACGCC TCCATGCTCG CATGGGTTCTT GGAAATATGCC GGCGCTGGGC
401 CGGGCTTCCT TATGGCGGC GTACC GGAA ATTTCCGCG TTTCGGCCG
451 CCTGCCGAA ACGCCGCGCC AGAACCGAA CAGCCATCG CGGTTTTG
501 TCTAGGAGC CGAGAATAC GACCGCGCT TttTCGACAA ACGTTCTAAA
551 TtCGGATT ACCTGGCGG TACCGCGCTG TTGACACATC TGAATTCTGA
601 CCACGGCAC ACCTGGCGG ACTTGGCGC GATACAGACG CAGTTCACT
651 ACCCTGTCG TACCGCGGC TCTGAAAGCT TAATGCTCTG CAACCGACCG
701 CAGCAAGCC TCGAAGATAC TTGGACAAAC GGCTGCTGGA CGCGGTGGA
751 AAAATTCGCG ACGGAAACAGC GTGCGCA..
```

This corresponds to the amino acid sequence <SEQ ID 866; ORF132>:

```

20      1 MKHIIHIIGIG GTFMGGGLAAI AKEAGFEVSG CDAKMYPPMS TQLEALGIDV
51 YEGFDQAQLD EFKADVVYVIG NVAKRGMDOV EA1LNLLPVL ISGPOLWSE
101 VLHHHWLVGV AGTHGKTTTA SMLAWVLEYA GLAPGFLLIG VPKGKFRRFP
151 PRAANAAPRPE QPIAVFRHRS RRIRHRLFQ TFXIRALPSA YRVEQSSGIR
201 PRRHLCLGR DTDFVFLPRA YRAVXRINRL QRTAARKPARY FGQRLLDAGG
251 KIRHGTRIA..
```

Further work revealed the complete nucleotide sequence <SEQ ID 867>:

```

1   ATGAAACACCA TCCATATTAT CGGTATCGGG GGCACGTTA TGGGCCGGCT
51  TGGCCGCATT GCAGAACGAG CGGGGTGTC AGTCAGCGGT TGGCACGGGA
101 AGATGATTCGC CGCCGATGAGC ACCCAAGCTCG AAGCCCTTGGG TATAGACGTG
151 TATGAAAGCT TGATGCGGC TCAGTGTGGAC GAATTTAAAG CGCACCTTTA
201 CGTTATCGGC ATATGCCCCA AGCGCGGAA CGGATGTTGGT GAAGCGGATTI
251 TGACCTCGG CCTGCCTAT ATTICCGGCC CGCAATGSGCT GTCGGAAAC
301 GTGCTGCAAC ATCATGGGT ACTCGTGTTG GCGGGGRCG ACGGCAAAAC
351 GACCCACGCC TCCATGCTCG CATGGGTTCTT GGAAATATGCC GGCGCTGGGC
401 CGGGCTTCCT TATGGGGCGC GTACCCGAAA ATTTCGGCTT TTCCGGCCGC
451 CTGCCCAAA CGGCCGCCAA AGACCGAAC AGCCATTCGC CGTTTTTCGT
501 CATGGAAGCC GACGAATACG ACACCGCTT TTTCGACAA CGTTCTAAAT
551 TCGTGCATA CGCGTCCCGT CGACCGCTGT TGAACANTCT GGAATTTCGAC
601 CAGCGGACA CCTTTGGCA CTGGGCCCG ATACAGACCC AGTTCCACTA
651 CCTCTGGCTT ACCTGGCGT CTGAGGCTT ATTCTGCTGC AAGCGACGCC
701 AACGAAAGCT CGGAAGATAC TTGGACAAAC GCTGCTGGC CGCGGTGGAA
751 AAATCGGC CGGAACACCG CTGGCGACCG CGCGAACGCC ATGCCGACGG
801 CTGTTGCGAC GTTGTGCTG ACGGCAAAAC CGCGGACGCC GTCAAATGGG
851 ATTGTGATGGG CAGCGCACAC CGCATGACAC CGCTCGCGT CATTGGCGGC
901 CGCGCTCATC TCGGTGTCGA TATTCAAGAC CGCTCGCAAG CCTTGGGGC
951 GTTAAACAAAC CGATGAAAGCC GGATGGAAACAT CAAAGGCAAG CGAAACAGGC
1001 TCACGGTTTA CGACGACTTC GCCCCAACAC CGACCGCCAT CGAAACACAG
1051 ATTCAAGGTG TGCGCCAAACG CGCCGGCGG CGGGCGCATCC TCGCGCTTCT
1101 CGAACCGGT TCCACACCGA TGAAAGCTGG CACGATGAAG TCGGCCTGTC
1151 CTGTAAGGCT CAAAGAACGC GACCAAGTGT TCTGCTACGC CGCGGGCGT
1201 GACTGGCGCT CGCCGNAAGC CCTCGGCCTT TTGGCGGCA CGCTGAACGT
1251 CGCGCAAGAC TTGATGCTT TCGTTGCGGA ATACGTGAAA AACGCCAG
1301 TAGGCGACCA TATTTGGTG ATGAGCAACG CGGGTTTCGG CGGAATACAC
1351 GGAAGCTGC TGGAAAGCTT GAGATAG
```

55 This corresponds to the amino acid sequence <SEQ ID 868; ORF132-1>:

```

1 MKHIIHIIGIG GTFMGGGLAAI AKEAGFEVSG CDAKMYPPMS TQLEALGIDV
51 YEGFDQAQLD EFKADVVYVIG NVAKRGMDOV EA1LNLLPVL ISGPOLWSE
101 VLHHHWLVGV AGTHGKTTTA SMLAWVLEYA GLAPGFLLIG VPKGKFRRFP
151 LPQTQDPDN EFKADVVYVIG DEYDTAFFFDR RSKFVHYRPF TAVLNNLEFD
201 HADIFADLGA IOTQFHVLVR TVPSEGILVC NGRQOSLQDT LDKGCTPVE
251 KFGTEHGWQA GEANADGSFD VLLDGKTAGR VKWDLMRHN MMNALAVIAA
301 ARHVGVDIQT ACEALGAFKN VKRMRMEIKGT ANGITYVDDF AHHPTRIETT
```

351 I QGLRQRWGG ARILAVLLEPR SNTMKLGTMK SALPVSLKEA DQVFCYAGGV
 401 DWDVAAELAP LGGRLNVGKD FDFAVVAEIVK NAEVGDHILV MSNNGFGGIH
 451 GKLLEAR*

Computer analysis of this amino acid sequence gave the following results:

5 Homology with the hypothetical o457 protein of *E.coli* (accession number U14003)

ORF132 and o457 show 58% aa identity in 140 aa overlap:

10	Orf132: 4	I HIIIGGTTFMGGGLAIAAKEAGFEVSGCDAKMYPFMSTQLEALGIDDVYEGFDAAQLDEFK	63
	o457: 3	I HIIIGCIGTFMGGGLAMLARQLGEHEVTGSDAANVPPMSTLLEKGCGIELIQQYDASQLEPQ	61
15	Orf132: 64	A D+V VIGNVAKRGMDDVVEAI LNRLGYYISGPQWLSENV LHHHHWLVGVAGTHGKTTTASL	123
	o457: 62	D+ VEA+L +FY+SGPQWL + VL WVL VAGTHGKTTTA M	121
20	Orf132: 124	AWLEYAAGLAPGFLIGGVXG 143	
	o457: 122	W+LE G PGE+IGGV G	141

Homology with a predicted ORF from *N.meningitidis* (strain A)

20 ORF132 shows 74.6% identity over a 189aa overlap with an ORF (ORF132a) from strain A of *N. meningitidis*:

	10	20	30	40	50	60
25	orf132.pep	MKHIHIIIGGTTFMGGGLAIAAKEAGFEVSGCDAKMYPFMSTQLEALGIDDVYEGFDAAQLD				
	orf132a	MKHIHIIIGGTTFMGGGLAIAAKEAGFEVSGCDAKMYPFMSTQLEALGIGVYEGFDAAQLD				
30	orf132.pep	10 20 30 40 50 60				
	orf132a	10 20 30 40 50 60				
35	orf132.pep	70 80 90 100 110 120				
	orf132a	70 80 90 100 110 120				
40	orf132.pep	130 140 150 160				
	orf132a	130 140 150 160 170				
45	orf132.pep	170 180 190 200 210 220				
	orf132a	180 190 200 210 220 230				

45 The complete length ORF132a nucleotide sequence <SEQ ID 869> is:

1 ATGAAACACA TCCACATTAT CGGTATCGGC GGCACGTTTA TGGGTGGGAT
 51 TCGGCCATT GCCAAAAGAG CAGGGTTGA ANTCAAGCGGT TCGCATGCCA
 101 AGATGATCGG GCGGATGAGC ACCGAGCTCG AAAGCTTGGG CATAGGGCTG
 151 TATGAAAGCT TCGCACCCGC GCAGTTGGG GAAATTAAAG CGCACGTTTA
 201 CGTTATCGGT AAATGTCGCCA AGCCGCGGGAT GGATGTGGTT GAAGCGATT
 251 TGAAACGGTG GCTGCTTAT ATTTCGGCT CGCAATGGCT GGCTGAAAC
 301 NTGTCGAC C ATCATGGNN ACTCGCGGC CGGGNGACGC ACGCCAAAC
 351 GACCCACCGG TCTATGCTCG CTGTGGTTT GGAATATGCC GGACTCGCAC
 401 CGGGCTTCTN TATCGCGGCC ATCGGGGGAA ATCTCAGCGGT TTCCGGCCCG
 451 CTGGCCCAA CGCCGCCGCA AGACCCGAAAC AGCCAAATGCC CGTTTTCTG
 501 CATTGAGGCC GCGAGATACG ACCAGCGGT TTTGAGCAAAAT CGCTCCAAAT
 551 TCGTGCATTCA CGTGGCGGT ACCGCCGGT TGAAACATCT GGAATTCGAC
 601 CAGCCGCCAA TCTTCGGCGA TTGAGGGGG ATACAGACCC AGTCCACCA
 651 CCTGGCTCGT ACCGGCGGT CTGAGGGCTT CATCGCTTGC AACGGACGCC
 701 AGCAAAGCTT GCAAGACACT TTGGACAAAG GCTGCTGGAC GCGGGTGGAA
 751 AAATTCGGCA CGGAAACAGCG CTGGCGAGGGC GCGGAAACCA ATGCGGATGG

	801	CTCGTTGAC	GTGTTGCTTG	ACGGCAAAA	AGCGGACAC	GTCGCTTGG
	851	GTTTGATGGG	CGGACACAAAC	CGCATGAAACG	CGCTCGCNGT	CATCGCCGCC
	901	GGCGCTCATG	CGGGACTGNGA	CATTGAGACG	GCCTGGCAAG	CCTTGAGCAC
	951	GTTTAAAAAC	GTCAAAACCCC	GCATGAAAT	CAAAGGCACG	GCAAAACGGTA
5	1001	TGACCGTTTA	CGACGACTTC	GGCCACCATC	CGACCCCTAT	CGAAACACAG
	1051	ATTCAAGGTT	TGGGCCAACG	CTTCGGGGGC	GGCGGCATTC	TGGCGCTCT
	1101	CGAACCCCGT	TCCAATACCG	TGAAGCTGGG	TACCATGAAA	GGCCGCGCTGC
	1151	CGCGAACGCT	CAAAGAACCC	GACCAAGTGT	TCTGNTACCG	CGCGCGGGCG
10	1201	GACTGGGACC	TTGGCGAACG	CTTCGGCCCT	TTGGCGGCCA	GCTGCACT
	1251	CGGCAAAAGC	TTGCGATGCC	TGCTGGCGGA	AATCTGAA	AAAGCCGAAG
	1301	CAGGGACCA	TATTGGGTG	ATGAGCAACG	CGGGTTTCGG	CGGAATACAC
	1351	ACCAAACCTG	TGGACGCTTT	GAGATAG		

This encodes a protein having amino acid sequence <SEQ ID 870>:

	1	MKHHIIH1IGIG	GTFMGGGIAAI	AKEAGFEXSG	CDAKMYPPMS	TQLEALGIV
	51	YEGFDATAQLD	EFKMADYVIG	VNAVKRGMDVV	EAILNRLGPY	ISGPQWLAEN
	101	ALHHHHWXLGV	ATHXHGTTTA	SMLAWLEYA	GLAPGFVIGG	VIFENFSVSAR
	151	LFOTPRQDPN	SQSFPFFVIAE	DEYDTAFFFDR	RSKFVHYRFR	TAVLNNELEFD
	201	HADI FADLGA	IQTQFHHHLVR	TVESEGLIVC	NGRQSLQDT	LDKGWTPTVE
	251	KFGTEHGWQA	GEANADGSFD	VLLDGKKRGAH	VANSIMGGH	RNMNALAVIAA
	301	ARHAGVDIOT	ACEALSTFKP	VKRRMEIKGT	ANGITVYDF	AHHPTAIEITT
	351	IQGLRQRVGG	ARILAVLPEP	SNTMKLGTMK	AALPASLKEA	DQVFXYAGGA
	401	DWDVAEALAF	LGGRLHVKGKD	FDAFVAEIVK	NAEAGDHILV	MSNNGFGGIH
	451	TKLLDALR*				

ORF132a and ORF132-1 show 93.9% identity in 458 aa overlap:

	25	orf132a.pep	MKHHIIH1IGIGGTFMGGGIAAI	AKEAGFEXSG	CDAKMYPPMS	TQLEALGIV
		orf132-1	MKHHIIH1IGIGGTFMGGGIAAI	AKEAGFEXSG	CDAKMYPPMS	TQLEALGIV
	30	orf132a.pep	EFKADVVYVIGNVAKRGM	DVVEA1INRGLPY	ISGPQWLAENX	LHHHHWXLGV
		orf132-1	EFKADVVYVIGNVAKRGM	DVVEA1INRGLPY	ISGPQWLAENX	LHHHHWXLGV
	35	orf132a.pep	SMLAWLVEYAGLAFGXF	IGGGVPENFSV	SARLPCQTPRQDPNSQSP	FFVIRADEYD
		orf132-1	SMLAWLVEYAGLAFGXF	IGGGVPENFSV	SARLPCQTPRQDPNSQSP	FFVIRADEYD
	40	orf132a.pep	RSKFVHYRFR	PFTAVLNNELEFD	HADI FADLGA	IQTQFHHLV
		orf132-1	RSKFVHYRFR	PFTAVLNNELEFD	HADI FADLGA	IQTQFHHLV
	45	orf132a.pep	ARHAGVDIOT	ACEALSTFK	NVKA	RRMEIKGT
		orf132-1	ARHAGVDIOT	ACEALSTFK	NVKA	RRMEIKGT
	50	orf132a.pep	ARILAVLPE	PRSN	TMTMKL	TMKRAALPASLKE
		orf132-1	ARILAVLPE	PRSN	TMKL	TMKRAALPASLKE
	55	orf132a.pep	FDAFVAEIV	VKNAA	EAGDH	ILVMSNNGFGGI
		orf132-1	FDAFVAEIV	VKNAA	EAGDH	ILVMSNNGFGGI

Homology with a predicted ORF from *N.gonorrhoeae*

ORF132 shows 89.6% identity over 259 aa overlap with a predicted ORF (ORF132ng) from *N. gonorrhoeae*:

60	orf132.pep	MKHHIIH1IGIGGTFMGGGIAAI	AKEAGFEXSG	CDAKMYPPMS	TQLEALGIV	YEGFDAAQLD	60
	orf132ng	MKHHIIH1IGIGGTFMGGGIAAI	AKEAGFV	SGCDAKMYPPMS	TQLEALGIV	HEGFDAAQL	60

-471-

	orf132.pep	EFKADVIVIGNVAKRGMDDVVEAILNLGLPYI SGPQWLSENVLRHHWLVGVAGTHGKTTA	120
	orf132ng	EFOADIYVIGNVAKRGMDDVVEAILNRLGLPYI SGPQWLSENVLRHHWLVGVAGTHGKTTA	120
5	orf132.pep	SMLAWLVEYAGLA PGFLIGGVXKGKFRFRP PAANAA PRPEQPIAVFVRHRSRR1RHLRFQ	180
	orf132ng	SMLAWLVEYAGLA PGFLIGGVXKGKFRFRP PAANAA PRPEQPIAVFVRHRSRR1RHLRFQ	180
10	orf132.pep	TFXIRALPSAYRVRVEQS GIRPRRHLCLRGRTDTPVPL RAYRAVXRVLNR LQRTAAK PARY	240
	orf132ng	T LQIRALPSAYRVRVEQS GIRPRRHLCLRGRTDTPVPL RAYRAVXRVLNR LQRTAAK PARY	240
	orf132.pep	FQGRLLDAGGKIRNGTRLA 259	
15	orf132ng	FQGRLLDAGGKIRNGTRLA 261	

An ORF132ng nucleotide sequence <SEQ ID 871> was predicted to encode a protein having amino acid sequence <SEQ ID 872>;

1 MKHIIHIGIG FTMGGIALL AKEFGAKVKG CKADMMPPMS TFOALGIV
20 HEGFDAAOLE EFGAQDIYVG NVAARRGMDVV FAINGLNRY ISGPQWLARN
51 VLIHHWLVGV AGTHGKTTT SMAWLVYRLL RFLGFLIGG PKVGFRRFLR
101 PTANASARPE QOIAVFRHRS RRIRHLRFQ TLQTRALSAF YVRREQSGINR
151 PRRHLRRLGID DTDPVPPFRS HRTIRRPHRL QRTAAKPEAY FGQLRLDAGG
201 KIHRTRRSLAD W*
251 KIHRTRRSLAD W*

Further work revealed the following gonococcal DNA sequence <SEQ ID 873>:

25	1	ATGAAACACA	TCCACATT	CGGTATCGC	GGCACGTTTA	TGGGGGGATAT
	51	TGGCCGCATT	GGCAAAAGAT	CCGGGTTAAC	AGTCAGCGGT	TGCGACGGC
	101	AGATGATCTC	GCGATGAGC	ACCAAGCTTC	GAATTCCTGG	CATAGGGCTA
	151	CACGAAAGCT	TGATGCGCC	GCAGTGTGGAA	GAATTCTTCAAG	CGATAATTAA
30	201	CGTCATCGG	AATGTCGCCA	GGCGGGGGAT	AGGATGGTGC	GAGGGCATTT
	251	TGAAAGGGC	GTCGCCATT	ATTTCCCGCC	CGCAATGCTC	GGCTTAAAAG
	301	GTGtcgcac	atccTTGgt	ATTCGGCGTC	ggcgaggac	ACGGGcaAAcc
	351	gaccsaCcGc	t-CATGCTCG	CTCTGGCTTC	GGAAATGTCG	GGCTGGCCG
	401	CGGGCTCTTC	CATCGGGGCT	tgaccggAAA	ATTTGGCTTC	TTCGGCCGG
	451	CTACCCACCA	GGCGCGGTC	AGACGGACAC	AGCAATGCTG	CGTTTTCGTC
35	501	CATGAAGACC	GACCAATAGC	ACACCGCTTC	TTTGACACAA	CGCTTCAAATG
	551	TCTGTGATTA	TGGCCGGCC	ACCGCCGTTG	TGAAACATTC	GAAGATTCGAC
	601	CAAGCGGACA	TCTTGGCGCA	CTGGGGCGG	ATACAGACCC	AGTTCCACCA
	651	CTCTGTGGCC	ACCGCTTACG	CGGAGGCTC	CATCGTCG	AGAACAGGAC
40	701	AGCAAAAGCT	CCGATGATCT	TTGACAAJAC	GTCTGTCGAC	GGCTGGTGGAA
	751	AAATTGCGCA	CGGGACACGG	CTGGCAGATT	GTCGAAGTC	ATGCCGAGGCG
	801	CTCTGTGAC	FTATTCCTTC	ACGGCAAAAA	AGCGGCCAC	GTCGTATGGC
	851	ATTGTGATTC	CGGAGCACAC	CGCATGACCC	CGCTGGCGT	CATCCCTGCG
	901	GCACGCGCAT	GGGAGGTGCA	TGTTGACAGC	GGCTGGGAG	CTTGTGGTGT
45	951	GTTTAAAGAC	GTCAACAGCC	GCATGAAAAAT	AAAGGGCAG	ACGAAAGCCG
	1001	ATTCAGCTTCA	CGAGCATTG	GGCACACACC	CGACCGCCAT	GAACCAACCG
	1051	ATTCAGGAGT	TGGCCCAACG	TGGCGGCGC	GGCGGATCAT	TCCGGCTTC
	1101	GGAGGGCGGT	TCTTACACCA	TGAAACTCGG	CACGATGAGC	TCGGCTCTGG
	1151	CGCGAACGCT	CAAAGAACGG	GACCAAGTGT	TCTGCTACG	CGGGGGCGGC
	1201	GACTGYGGACG	TGGCCGACG	CTCTGGCGCT	TTGGGGCTGA	GGTCGGCGG
	1251	CGGTAAGAGT	TTGCAATACCT	TCTGGCCGA	ATTGTTGAA	AAAGCCGGAA
	1301	CGGGCGACCA	TATTGTTGGT	ATGAGCAACG	GGGTTTCG	CGGAATACAA
	1351	BCCACATCG	TGGGAGCTT	GAGATAG		

This corresponds to the amino acid sequence <SEQ ID 874: ORF132ng-1>:

55 1 MKHHIIHIGG GTFMGGIIA1 AKEAGFKVSG CDAKMYPPMS TQEALGIVG
51 HEGFDAAOLE EFOADYIVVG NVARRMDVV FAEILNGFLY ISFGQFWLA
101 VLLHHWWLGE AGTHGKTTA SMLAWVLEYA GLAFGLIGI VFENGVGSAR
151 LPCTQPRDNG SKSSPFVVE DEYDTFFDK RSKFVHYRPR TAVLNLEFNE
201 HADIFADIGA IOTOFHHLRV TVPSEGLIVC NGQOQLSLOQT LUKGWPVPE
251 KFGTGHGWII GEVNADGSDF VLLDGKGKH VAWDLMMGHN RMNLAVIAA
301 ARHAGWVQD ACEALGAFKN VKRMRHEIKGT ANGITVUDDE AHHTFAIZDE
351 IOGLQRGRVGG ALAFLVLEER SNTMKLTGMK SASPALESKA DCVQFYAGGA
401 DWVAEALAP GCLCRRLVRGKD FDTFVAEIVK NARTGSHILV MSNNGGGIIN
451 TKLILDA*
501

ORF132ng-1 and ORF132-1 show 93.2% identity in 458 aa overlap:

```

5      orf132ng-1.pep MKHIIHIIGIGGTFMGGIAAIAKEAGFKVSGCDAKMYPPNSTQLEALGIGVHEGFDAQLE
      |||:|||||:|||:|||||:|||:|||||:|||:|||||:|||:|||||:|||:|||||:|||:|||||:|||:|||:|
orf132-1      MKHIIHIIGIGGTFMGGIAAIAKEAGFEVSGCDAKMYPPNSTQLEALGIDVYEGFDAQLO
10     orf132ng-1.pep EFQADIIYVIGNVARQGMDVVEAILNRGLPYISGPQWLAEVNVLHHHWVLGVAGTHGTTTA
      |||:|||||:|||:|||||:|||:|||||:|||:|||||:|||:|||||:|||:|||||:|||:|||||:|||:|||:|
orf132-1      EFKADIVVIGNVARQGMDVVEAILNLGLPYISGPQWLSENVLHHHWVLGVAGTHGTTTA
15     orf132ng-1.pep SMLAWVLEYAGLAPGLFGLGGVPENFGVSARLPQTPRQDPNSKSPFFFVIEADEYDTAFFDK
      |||:|||||:|||:|||||:|||:|||||:|||:|||||:|||:|||||:|||:|||||:|||:|||||:|||:|||:|
orf132-1      SMLAWVLEYAGLAPGLFGLGGVPENFGVSARLPQTPRQDPNSKSPFFFVIEADEYDTAFFDK
20     orf132ng-1.pep RSKFVHYRPTAVLNLLNEFDHADIFADLGAITQTFHHHLVRVVSSEGLIVCNGQQQLSDT
      |||:|||||:|||:|||||:|||:|||||:|||:|||||:|||:|||||:|||:|||||:|||:|||||:|||:|||:|
orf132-1      RSKFVHYRPTAVLNLLNEFDHADIFADLGAITQTFHYLVRVVSSEGLIVCNGRQQLSDT
25     orf132ng-1.pep LDKGCGWTPVERFGHGWQIGEVNAIGSFDFVLLDGKKAGHVANDLNGHHNRMNALAVIAA
      |||:|||||:|||:|||||:|||:|||||:|||:|||||:|||:|||||:|||:|||||:|||:|||||:|||:|||:|
orf132-1      LDKGCGWTPVERFGHGWQIGEVNAIGSFDFVLLDGKTAGRVKDLMGRHNRMNALAVIAA
30     orf132ng-1.pep ARHAGVDVDQPACEALGAFKNVKRRMEIKGTANGITVYDDFAAHHTAIEETTIQGLRQRVGG
      |||:|||||:|||:|||||:|||:|||||:|||:|||||:|||:|||||:|||:|||||:|||:|||||:|||:|||:|
orf132-1      ARHVGVDIQPACEALGAFKNVKRRMEIKGTANGITVYDDFAAHHTAIEETTIQGLRQRVGG
35     orf132ng-1.pep ARIALAVLEPRSNNTKLGTMKSALPSLKEADQFCYAGGAQDWVAEALAPLGCRLRVGKD
      |||:|||||:|||:|||||:|||:|||||:|||:|||||:|||:|||||:|||:|||||:|||:|||:|
orf132-1      ARIALAVLEPRSNNTKLGTMKSALPSLKEADQFCYAGGAQDWVAEALAPLGCRLRVGKD
40     orf132ng-1.pep FDTFVVAEIVKNARTGDHIILVMSNGGGGGIHTKLLDALRX
      |||:|||||:|||:|||||:|||:|||||:|||:|||||:|||:|||||:|||:|||:|
orf132-1      FDTFVVAEIVKNAGVGDHILVMSNGGGGGIHTKLLDALRX

```

In addition, ORF132ng-1 is homologous to a hypothetical *E.coli* protein:

```

35     pir|IS56459 hypothetical protein o457 - Escherichia coli >gi|537075 (U14003)
orf o457 [Escherichia coli] >gi|1790680 (AE000494) hypothetical 48.5 kD protein
in fbp-pmba intergenic region [Escherichia coli] Length = 457
Score = 474 bits (1207), Expect = e-133
Identities = 249/439 (56%), Positives = 294/439 (66%), Gaps = 13/439 (2%)
40     Query: 22 KEAGFKVSGCDAKMYPPNSTQLEALGIGVHEGFDAQLEEFQADIIYVIGNVARQGMDVVE 81
++ G +V+ DA ++YPNST LE GI + +G+DA+QLE P Q D+ +IGN RG VE
Sbjct: 21 RQLGHEVPGSDANVYPPNSTLEEKQGIELBQGYDASQLEP-QPDVLIIGNAMTRGNPVE 79
45     Query: 82 AILNRGLPYISGPQWLAEVNVLHHHWVLGVAGTHGTTTASMLAWYAGLAPGLFGLGGV 141
A+L +F+ SGPQWL + VLV WVL VAGTHGTTTA M W-LG P PGF+IGGV
Sbjct: 80 AVLEKNIPYMSGPQWLHDVFVLRDRWVLVAGTHGTTTAGMATWILEQCGYKPGFVIGGV 139
50     Query: 142 PENFGVSARLPQTPRQDPNSKSPFFFVIEADEYDFAFDKRSKFVHYRPTAVLNLLFDH 201
P NF VS L A +S FFVIEADEYDFAFDKRSKFVHYRPTAVLNLLFDH 201
Sbjct: 140 PGNFVEVAHL-----GESDFFVIEADEYDCAFDKRSKFVHYCPRTLILNNLEFDH 190
55     Query: 202 ADIFADLGAITQFHHLVRVTVPSSEGGLIVCNGQQQLSDTLDKGCGWTPVERFGHGWQIG 261
ADIF DL AIQ QFHHLVR VP +G I+ +L+ T+ GCW+ E G WQ
Sbjct: 191 ADIFADLKAQKQFHHLVRIVPGQGRIIWNPENDINLKQTMANGCWSEQLVGEQGHQAK 250
60     Query: 262 EVNADGS-FDVLLDGKKAHVANDLNGHHNRMNALAVIAARHAGVDVDQFACEALGAFKN 320
++ D S ++VLLDG+K G V W L+G HN N L IAAARH GV A ALG+F N
Sbjct: 251 KLTTDASEWEVLLDGKEVGEVKWSLVGEHHNNGLMAAARHVGWAPADAANALGSFVIN 310
65     Query: 321 VKRRMIKGTAGTANGITYDDFAAHHTAIEETTIQGLRQRVGG-ARIALAVLEPRSNNTKLGTM 379
++ RR+E+G ANG+TVYDDFAAHHTA T- LE +VGG ARI-AVLEPRSNNTKMG+G
Sbjct: 311 ARRLELPYERGANGVTVYDDFAAHHTATLAALRGVGGTARIIAVLEPRSNNTKMGIC 370
70     Query: 380 KSALPASLKEADQVF-CYAGGAQDWVAEALAPLGCRLRVGKDFTFVVAEIVKNARTGDHI 438
K L SL AD+VF W VAE D DT +VKG A+ GDHI
Sbjct: 371 KDDPLASLGRADEFLQPQAHIPWQVAEVAEACVQPAHWSGVDLADMVVTKAQPFGDHI 430
75     Query: 439 LVMSNGGGGGIHTKLLDAL 457

```

LVMSNGGFGGIH KLLD L
Sbjct: 431 LVMSNGGFGGIHQKLLDGL 449

Based on this analysis, it was predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

- 5 ORF132-1 (26.4kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 20A shows the results of affinity purification of the His-fusion protein, and Figure 20B shows the results of expression of the GST-fusion in *E.coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for FACS analysis (Figure 20C) and ELISA (positive result). These 10 experiments confirm that ORF132 is a surface-exposed protein, and that it is a useful immunogen.

Example 103

The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 875>

	1	..CCGGCTTATT	ACGGCTCGGA	TGACGATT	AACGGGGCAT	TCGGGAAA
15	51	CTCGGCCGAA	TACAGAAC	ATTCACGAA	GCGCTACGGG	TTATTCGACG
	101	CGGTATTGGA	AAATAACG	AAAGGGCGG	CCAAACCA	TCCTGGCACA
	151	ATTAGTCCG	ACTTGCGGA	TTTACCGG	CCTGGTGC	GATTTACGGG
	201	CACACACCT	ATGCCCAA	TCCRAAAT	TTTATTTCCA	AAATCCGGC
	251	ACTCCGGCT	TCACACCC	TTAAACCG	AGCGGCCG	ACTTGGCC
	301	TTTGGCTTC	ATACATCA	AAAGGATTG	TTAACACAG	ATGATACAT
20	351	AGGATAAA	CTGTCGGCT	ACCCCGAC	CATTCGACAC	TAATCCACCA
	401	ACGTTTACGG	GAATATGG	GATTTCGACG	GGGATATTCC	GAGCTGGTC
	451	AGCACGACC	GGCTTCGCTA	CACCATCC	CATCCGATT	TAACAGACAA
	501	AGTCGATCAA	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	TTATGGGTT
	551	TTTCACCAA	CCTTCTTAC	GCCTATCAA	AAAGCACC	ACCGGACCAA
25	601	TTTACGGATG	CGACGGGATC	CCCCAACAT	GCTCTTCAA	AAAGAACCT
	651	CAACAAAGG	TATGGTTG	CGAAGGGT	CGCCGCGG	CGAGGAACT
	701	GCCTTTGG	ACTGGCTG	CCCTGTTG	CGGCGGCGG	GATTTGGC
	751	GGCGGGATGC	GTATTTGG	CAAGAGCATC	CGGGCGACGG	CTGAGAACG
	801	CTATATCGAC	GGCCACAAAG	GGGGAAAT	CAGCAATTTC	GCTTTGGCG
	851	CGAACGCTTC	CTAACACAA	ACCGGAAAC	TTGGCCGCA	CGCAACTGG
	901	TTWAGTTTT	ACCCGGCTT	CGACGGCG	AAAACCTTA	TTTCCGGG
	951	CGAGCTCAA	ATACCTTC	ACAGGGCTTA	TATCGATC	TCAGATCGG
	1001	GGCATGATG	GGCAGCA	GAGCTTATTA	GCTGTG	CCGGAAAGAC
	1051	AAGAGC	ATGACCTAC	TTATGCTG	AAACAGCTG	TACAGCGCA
35	1101	ATAGCGGAT	ACGAACTAA	GGCTGATGG	CAATTTCG	CGCGGACGCA
	1151	CTTTGGAT	GGACAGGAGC	TAACAGTTT	AA	

This corresponds to the amino acid sequence <SEQ ID 876; ORF133>:

1 . . . PGYGSODEP KRAFTGENPT XKRKHGSQSS YIEPKLWKKY KRANHNSVS
51 ISADFGDYFM PFSYASRTHR MPNQJCEMYS QIGDGSVHTA KERPLARANTO
101 FCGTYXIKGKL LKQDUDLGLR LVLYGRSXRQI YIHNHYGKWRW DLNDGJLWTRT
151 SSTGLATIQI HRFXDXWHRQ XXXXXXXXXD YGRFLFTNLSY AQXKSTOPNTN
201 FSDASESPNN ASKEDDLQKG YGLRSVRSLAP RDYLGRLVEGT WLNLGNKLRT
251 GMARYFKSLS RATAEERYID GTNGNSHNF RQLGRKSISQ TLQRZLQLP
301 XDPAHAYEPK KNLFLPRAYWD NLDFDRYLT LDAGNDAAXE RYSSFDPKD
351 KDNVCCBBD TCKLNGKCYGG TSFKSWLNFPA EBLFELMTMS YKF

Further work revealed the further partial DNA sequence <SEO ID 877>:

1 GAGGCCGCGA TACAGCTT GGAGATGTG CACCTCAAGG CGAGCCGGC
 51 ACCGAAAGAC AAAAACAGTG TTACCGTCG CGGTGCGCA TGCGACCGTC
 101 AGGATATTG AAAAACTTCG ACAAATCTCG AACAATCTCG AGCACAGCAT
 151 CCCGGTGCCT TTACAGCGCA AGATAAANG CGGGCGATG TCTTCTTGA
 201 TATTCCGGG GACAGCGCC TCGGGCGGTG AAATAGGATG GTGGAAGGCA
 251 TCACGGGAG CTTTATTTCG ACTTCTACCG ATGGCGGCA CGCCAGGGC

301	TGATCCTCAAT	TCGGTGATC	TGTGCGACAGC	AATTTTTATTG	CCGGACTGGA
351	TGTCGCTAAA	GCGGACTTCA	GGCGCTCGGG	AGGCATCAAC	AGCCCTGCGG
401	GTTGGGGAA	TCTGCGGACT	TTAGGCGTGG	ATGACGTCTG	TCAAGGGCAT
451	AATACCTAACG	GCGCTGCTCT	AAAAGGTCTG	ACCGGACCCA	ATTCAACCAA
501	AGGTAAATGCC	ATGGCGCGG	TAGGTCGGG	CAAATGCGT	GAAAGCGGAG
551	CATCTGCGG	TGTCGCTTAC	GGGCACAGCA	GGCCGAGCGT	GGGGAAAAT
601	TACCGCTG	GGGGCGCGG	CCACACATC	GGAAATTTTG	GGCCGGAATA
651	TTTGGAAACGG	CGCAACAGC	GATATTTTGT	ACAAGGGGT	GCTTTGAAT
701	TCATATCGG	CAGGGAAA	TGGGAGGGG	ATTTCACAA	GCACAGTGG
751	AAATACACG	CUTATAAAAA	TTAACACAM	CAAGAACTAC	AaaATACAT
801	CGAAGACAT	GACAAACACT	GGCGGGAAA	CCTg.CaCC	CAATACGACA
851	TTTACCCCAT	GTGATCGTC	AGCTTGAGGA	AGCAGTCGGG	AGGCAATCTG
901	TTTAATATTG	AAATCACCGG	CTGATTCAT	AAATACACG	CCCAATTTCG
951	CGATTTAAC	ACCCAAATCG	GCAGCGCGGA	AAATCAACG	CCCAATTATC
1001	AGTTCATTA	CGGTGTCCT	TGGAACCGT	ATACCAACCT	CAATCTGACC
1051	CGACGCTACA	ATTCGGGAG	CGAACATAT	CGGAAAGGGT	CGAAGTTATC
1101	AGGCTGGGG	TTTCTTAAAG	ATTTGAAAC	CTAACACAA	GGGAAATCC
1151	TCGACTCTCA	CAACACCGC	ACCTTCGGG	TGCCCCGGGA	AACCGAGTTG
1201	CAACACCT	TGGGCTTCA	TTATTTCCAC	AAACAAATAG	GCAAAAACCG
1251	CTTCTTCAAG	GAATTGGGC	TGTTTTTGG	CGGTCTGAT	CAGCAACCG
1301	CGCTTATTC	CTATTTGGG	CGCTTAAAG	GGCATTAAGC	CTGCTGCC
1351	CAAAATCAA	CCATGTC	ACCGCGCGC	AGCCAAATT	TCAACAGCTT
1401	CTACTCGAT	GCGCGCTCA	AAAAGACAT	TTACCGCTT	AACATACAGCA
1451	CCAAATACG	CGGCTACCGT	TTGGCGCGG	AAATACACGG	CTATTCACGC
1501	TCGGATGAGC	ATTATAGGC	GATCTGGG	GGAAACTCTGG	CGACATACAA
1551	GAACACCT	GGCGGCAAC	GGGGGATTA	TGACCCCGT	TGAAAAAAAT
1601	ACGGAAAAAA	GGCGGCAAC	AACCATTCGG	TCACCATTTAG	TGGGACTTTC
1651	GGCGATTATT	TCATGGCTT	CGCCAGCTAT	TGGCCGACAC	ACCGTATGCC
1701	CAACATCCAA	GAATGATTA	TTTCGAAAT	CGGGGACTCTC	GGGTTTACA
1751	CGCGCTTAAA	ACCAAGGGC	GCAAAACACT	GGCAATTG	CTTCATATCC
1801	TATAAAAG	GATTTGTTAA	CAAGAATGAT	ACATTAGGAT	TAAAATCTGT
1851	CGGCTTACCGC	AGCCGCTAC	AAACTACAT	CCACAACTGT	TACGGGAAAT
1901	GGTGGGATTG	GAACGGGAT	ATTCGGAGGT	GGGTCAAGCAG	CACCGGGCTT
1951	GCTTACACCA	TCCAACATG	CMATTTCAAA	GACAAAGTCG	ACAAAACCGG
2001	TTTGAGTTG	GAGCTGAAT	ACGGATTATGG	CGGTTTTTC	ACCAACCTTT
2051	CTTACCGGCA	TTTAAACGG	ACCCACCGA	CCAACTTCAG	CGTGGCGGAC
2101	GAATCGCCA	ACAATGCTC	CAAAGAACAC	CAACTAAAC	AAAGTTATGG
2151	GTGAGCAGG	GTTCGCGCC	TGCGCGAGA	TTACGGACGT	TTGGAAGTCG
2201	TTACGGCGCT	GTGGGCAAC	AAACTGACTT	TGGCGCGC	GATGGGTAT
2251	TTOGCGAAGA	GCATCGCG	GACGGCTGAA	GAACGCTATA	TCGAGGCGAC
2301	CRACGGGAA	AAATACCGA	ATTTCGCGA	ACTGGCGAAG	CGTTCATCA
2351	AAACAAACCGA	PACTTCTGC	CGCCAGCGCT	TGATTTTTG	TTTTAACGCC
2401	GCTTACGAGC	CGAAGAAA	CCTTATTTC	CGGCCGAA	TCAAAATCT
2451	GTTCGACAGG	CGTTTATCG	ATCGCGCTCGA	TGGGGCGCAT	GTAGGGCGAA
2501	CCCGCTTAA	TTACACGTC	TTGGACCGGA	AAGACAGGA	CGAGACGTA
2551	ACGTATATG	CTGATAAAC	GTGGCAAC	GGCAAATAC	GGGACACAAAG
2601	CAAAGCGTA	TTGACCAATT	TTGACCGGG	ACGCACCTT	TGATGACGA
2651	TGAGCTACAA	TTTTTA			

This corresponds to the amino acid sequence <SEQ ID 878; ORF133-1>:

50	1	EAQIQVLEDV	HVKAKRVPKD	KVVFDTARAV	STRQDFIKSS	ENLDNIVRSI
	51	PGATPQKDS	SGIVSLLNRG	DSPGRWNTM	DVGITQTFYS	TSTDAGRAGG
101	SSQFGASVSDS	NFIAGLDDVV	GSPSGSAGIN	SLAGSANLRT	LGDDVVQGN	
151	NTYGLLRLGL	TGTNSTKGNA	MAIGARKWL	ESGASGVLYI	GHSRSRVAQN	
201	YRVGGGQHNI	GPNFGEAYLRR	RKQRFYVQEG	ALKENSDSGK	WERDLQRQW	
251	KVFKPYNNN	CELKQKVIIEH	DKSWRENLXQ	QDIDTFIDPS	SLKQQSAGNL	
301	FKLEYDGVN	KTYAQFRDIN	TKIGSRKIIN	RNYQFNQYLS	LNFYTNLNLT	
351	AAYNNSGRQKY	PKGSKETKGW	LKKDFETYNN	AKILDLNNTT	TFLFRETEL	
401	QTTLGFLNYFH	NEYKGKNRFP	ELCLPFDGP	QDNLGYSYLQ	RFKGDKGLLP	
451	KGSTLIVQPG	SQYENPFYFD	AALKKDIYRL	NYSTNTVGYR	FGEGETGYG	
501	SDDEPKRFAE	ENSTPYKHKD	NSSCCIYEVP	LKKYKGKRN	NHSVSISGADF	
551	GDFMFPMASY	SRTHRMPNIQ	EMYFSQIGDS	GVHTALKPER	ANTWQFGENT	
601	YKKGLLKQDD	TGLKLKVGYR	SSLDINYIHN	YKGKWDLNGD	IPSWWSSTGL	
651	AYTIQHNRKE	DVKVHKIGFEL	ELNDYDGRPF	TNLSIYAYQKS	TQPTNFSDAS	
701	ESBNNAQSKED	OIKQGVLGLSR	VGAFLPDGVE	LEVGTBWLGN	KLTLLGGMARY	
751	FGKSIIRATAE	EYRIDGTNGG	NTSNFRQLGK	RSIKQETLIA	RQPLIFDFYA	
801	AYPEPKNLF	RAEVKNLDR	RYIDPLDAGN	DAATQRYYSS	FDFPKDKDEDV	
851	TCAKDRTLCN	GKYGGTSKSV	LMTFARGKTE	LMTMSYK*		

Computer analysis of this amino acid sequence gave the following results:

Homology with the probable TonB-dependent receptor HI121 of *H.influenzae* (accession number U32801)

ORF133 and HI121 show 57% aa identity in 363aa overlap:

Orf133: 31 IYEPVULKYKKRANNESVSISADFGDYMPFPASYSRTHRMPNQIEMYFSQIGDSCGVHTA 90
 5 I EP+L K G K+A NES ++SAV DYPMPF +YSRTHRMPNQIEMYFSQIGDSCGVHTA 90
 HI121: 563 INEPLIKQGHHKKAFAFNESATLSAELSDFMPPFTYSGTHRMPNQIEMYFSQVSAGWNTA 622

Orf133: 91 LKPERANTWQFGFXTYKKGLLKQDDTLGLKLVLGYRSRIDNYIHNVYGKWWDLNGDIPSPWV 150
 10 LKEP+T+T+Q GF TYRKGL QDD LG-KLVGYRS VD YIHNHYVG WW -P+H
 HI121: 623 LKPEQSOTYOLGFNTYKKGLFTQDDVLGVKLVLGYRSFIRKNYIHNVYGVW--RDGMPTWA 680

Orf133: 151 STSGLAYTQIHRXFKDVHQXXXXXXYDGRTFTNLSSAYQKSTQPTNFSDASESPNW 210
 15 S G YTI H+ + V YD GRFF N+SYAQ++ QPTN++DAS FNN
 HI121: 681 ESGNFGKYTTIAHONYKPFVKKSGVLEINMDGRFVFANVSYAYQRTNQPTNYYADASPRFNN 740

Orf133: 211 ASADEQLQKGYGLSRVSLP+DRGLRLEVGTRWLGNKLTLGGAMRYFGKSIRATAEERYID 270
 20 AS+ED LKGQYGLSRVS LP+DRGLRLEVGTRWLGNKLTLGGAMRYFGKSIRATAEERYID 270
 HI121: 741 ASQEDILKQKGYGLSRVSMPLKDYGRLLEGLTRWFQDKLTLGLAARYYGKSKRATIEEEYIN 800

Orf133: 271 GTNGGNTSNFRQLGKRSIKQTETLARQPLIXDGFNAAYEPKKNLIFRAEVKNLFDRRYIDP 330
 25 G+ + R+ +K+TE + +QPI+D + +EP K+LT +AEV+NL D+RY+DP
 HI121: 801 GSR-FKKNTLRRENYVAVRKTSEDIKKQFIIQLDHLVSYEPIKDLIKAEVQNLLDKRYVBP 859

Orf133: 331 LDAGNDAAKERYYSSFDPKDKDVTCAKDTLCLNGKYGGSKSVLTNFARGRFLMTMS 390
 30 LDAGNDAA +RYYSS + C D + C GG+ K+V L NFARGRY++++++
 HI121: 860 LDAGNDAAQSRYSSL----NNSTIECAQDSSAC---GGSDKTVLYNFARGRYIILSLN 910

Orf133: 391 YKF 393
 YKF
 HI121: 911 YKF 913

Homology with a predicted ORF from *N.meningitidis* (strain A)ORF133 shows 90.8% identity over a 392aa overlap with an ORF (ORF133a) from strain A of *N.meningitidis*:

35	orf133.pep	10 20 30	
		PGYYGSDEEFKRAFGENSTTXKKHCNRSCGI	
	orf133a	: : : : :	
	PYFDAALKDIYRLNYSINTVGYRFGGYTGYIYXSDEEFKRAFGENSTYKHCNCQSGI	450 460 470 480 490 500	
40	orf133.pep	40 50 60 70 80 90	
		: : : : :	
	orf133a	YEFPVULKYKKRANNESVSISADFGDYMPFPASYSRTHRMPNQIEMYFSQIGDSCGVHTA	510 520 530 540 550 560
45	orf133.pep	100 110 120 130 140 150	
		: : : : :	
	orf133a	YEPFVULKYKKRANNESVSISADFGDYMPFPASYSRTHRMPNQIEMYFSQIGDSCGVHTA	570 580 590 600 610 620
50	orf133.pep	160 170 180 190 200 210	
		: : : : :	
	orf133a	STGLAYTQIHRXFKDVHQXXXXXXYDGRTFTNLSSAYQKSTQPTNFSDASESPNW	630 640 650 660 670 680
55	orf133.pep	220 230 240 250 260 270	
		: : : : :	
	orf133a	SKEDQLKQGGLSRVSLPDRYGRLEVGTRWLGNKLTLGGAMRYFGKSIRATAEERYIDG	690 700 710 720 730 740
60	orf133.pep	280 290 300 310 320 330	
		: : : : :	
	orf133a	SKEDQLKQGGLSRVSLPDRYGRLEVGTRWLGNKLTLGGAMRYFGKSIRATAEERYIDG	690 700 710 720 730 740
65	orf133.pep	TNNGGNTSNFRQLGKRSIKQTETLARQPLIXDGFNAAYEPKKNLIFRAEVKNLFDRRYIDP	

orf133a	 TNGXXTSNFRQLGKRSIXQETELARQPLIFDXYAAYPEKKXLIFRAEVKNLFDRRYIDPL 750 760 770 780 790 800
5 orf133.pep	340 350 360 370 380 390 : : DAGNDAAXERYYSFDPKDKDXDVTNCNDTLCNGKYGGTSKVLTNFARGRTFLIMTMSY
orf133a	: DAGNDAATQRYYSSFDPKDKDEEVTNCNDTLCNGKYGGTSKVLTNFARGXTFLITMSY 810 820 830 840 850 860
10 orf133.pep	KFX
15 orf133a	KFX 870

A partial ORF133a nucleotide sequence <SEQ ID 879> is:

```

1 AAAGACAAAA AAGTGTTCAC CGATGCGGT GCGGTATCGA CCGTCAGGA
51 TATATTCAA TCCANCGAAA ACUTCGACAA CATCGTACGC ANCATCCCC
101 GTGCGTTAAC ACANCAANAT AAAAGCTCGG GCNITGTGTC TTGAATATT
151 CGCNCGACAA CGCGGTTGCG GCGGGTCAAT ACNATGTCNA ACGGCATCAC
201 NCANACCTTTC TAFTCGCTTC TAFCAGATGCG GGGCAGGGCA GCGGTTTCAT
251 CTCAATTGCG TGCAATGTCG GACAGAAATT TTATNGCGG ACTTGGATGTC
301 GTCAAAAGGCA GGTTCAGGGG CTCCGGAGGG ATCAACAGGC TTGGCGGTTTC
351 CGGCAATTCG CCGGATTTAAGG GCGTGGATGTA TGCGTTCAG GCGAATANTA
401 CNTACGGGCTT CGFCTTAAGG GGTCTGACCC GCACCAATTTC AACCAAAGGT
451 ANTSGCGATGG CGCGGATAGG PGCCGCAAAA TGCGTGGAAA GCGGAGGCATC
501 TGTCGGGTTGCTT CAACAGGGC ACACAGGGC CAGCGTGGCG CAAATTACCC
551 CGGTGGGGCG CGGGGGGAGG ACATCGGAA ATTTCGGCGG GGAAATATCTG
601 GAACGACGCA AGCAACGATA TTTTGACGAA GAAGGGGGGT TGAATTTCAA
651 TTCCAAACAGC GGAAATGGG AGCGGGATTTT CCAAAAGTCG TACTGCGAAA
701 CCAAGTGTAA TCAAAATATC GATGCCCGCC AAGAACTGCA AAAATACATC
751 GAAGGTCTAT ATAAAAGGCTT GGGGGAAAAC CTGGCGCCGG AATAGACGACAT
801 CACCCCGATC GATCGCTGCA CGCTGAGCN GCAGTCGCG GCGAACCTGT
851 TTAAATGGAA ATACAGGGC GTATTCATAA AATACACGGC GCAATTTCGC
901 GATTAAACA CCAAAATTCG CAGCGGAAA ATCATCAACC GCAAAATATCA
951 AITCAAITAC GGTGTTCTCTP TGAACUGGIA TACCAACCTTC ATATCTGACCC
1001 CGACCTACAA CGCGGCGG CAAAGATATC CGAAAGGGT GAGTTTACA
1051 GGCTGGGGCC TTTTNAAGG TTTGGAAACCT TCAACAAACG CAAAATTCCT
1101 CGACCTCAGC AACACTCTCA CCTTCGGGT GCGCCGTGAA ACCGGATGTC
40     1151 AAACCACTT GGCCTTCAT TATTTCCCAA ACGAATACGGG CAAAACACCGC
1201 TTTCCTGGAAAGG TTTGGGGCTT GTTTTGCGAC GGTTCGGATC ANGACAACCGG
1251 GCTTTATTCG TATTGGGGC GTTTTAAGGG CGATAAAGGG CGTCTGCCCC
1301 AAAAATCAAC CGATTCGCA CGGCCCGCA GCGAAATATTTC CAACACCTTC
1351 TACTTCGATG CGCGCGCTAA AAAGACATI TACCGCTTAA ACTACAGCAC
1401 CAATACCGTC CGGCTACCTTC TGCGCGCNA ATATACGGG TATTACNGCT
1451 CGGATGAGCA ATTAAAGGGG GCGATTCGGG AAAACTCCGC GICATACANG
1501 AAAATCTGCA ACCAGGCTGCG CGGAATTAT GAACCGGTAT TGAAAAATTA
1551 CGGCAAAAGG CGGGCCAAACA ACCAATTCGGT CAGCATTAGT CGGGAATCTG
1601 CGGATATTTCGATCGCGTTC CGCAGCTTAA CGGCACACAA CGTATTCGCC
1651 AACATCCAAAG AATTTGTTT TTCCCAAAATC GGGGACTTCG GGGTTCACAC
1701 CGCCCTTAAAGA CGACAGGGG CAAACACTTG GCAATTGGG TTCAATACCT
1751 ATAAAANAAAAGG TTGTTAAAGG CAAGATGATA TATTAGGATT AAAACTGGTC
1801 GGCTACCGCA CGCGCGATCGA CNACTACATC CACAACGTTT ACGGGAAATG
1851 GTGGGATTTCG AACGGGAAATA TTCCCGACTG GGTGAGCAGC ACCGGGCTTG
1901 CCTACACCATC CAACACCGG AATTTCAAGG ACAAAGTCGA CAACACACGGT
1951 TTGAGTTGG AGCTGATTA CGATTATNGG CGTTTTTC CAACACCTTC
2001 TTACGGCTTAA CAAATGGCGG CGCAACCGAC GAACTTCAGC GATGGGAGCG
2051 AATCGCCCAA CAATCGCTG AAAAGAGACCA AACTTCACAA AGGTTATGGG
2101 TTGAGCAGGG TTTCGCGCCC GCGCGGAGAT TACGGACGTT TGGAAGTCGG
2151 TACGGCTGG TTGGGCAACA AACTGACTTTT GGGGGCGGT ATGCCCTATT
2201 CGAGCGTTATC FTAGCTGGT CGCGCGGAGG AACCGCTTAT CAGACGCCAC
2251 AATGGGNNAN TTACAGGCAAGG TTCCGGCGAA CTGGCGAAC GTTCCCATCAN
2301 ACAAAACGGCA ACCCTTGGCC GCGACGGCTT GATTTTGTAT TTNTACGGCG
2351 CTTACGCGGAA GGAAGAAAATC CTATTTTCG CGCCCGAAGT CAAAATCTG
2401 TTGCGACAGGG GTTATATCGA CGCTCTGAT CGGGCGATG ATGCGCGAAC
2451 CGACGGTTATC FTAGCTGGT CGCGCGGAGG AGACAAAGGAC GAAGAAGTAA
2501 CGTGTAAATG TGATAACACG TTATGCAACG GCAAAATACGG CGGCACAAAGC
2551 AAAAGCGTAT TGACCAATT TGCAACGGG CGACCGCTTT TGATAACGAT
2601 GAGCTACAAAG TTTTAA

```

This encodes a protein having (partial) amino acid sequence <SEQ ID 880>:

	1	KDKKVFTDAR AVSTRQDIFK SXENLDNIVR XIPGAFTXQX KSSGVSVLSN
	51	RXDSGEGRVN TMVDGITXF YSTSTDAGRA GSSSQFGASV DSNFXAGLDV
5	101	VKGSTFSAGC INSLAGSANL RTLKVDWDVO GNXTYGLLLK GLTGTNSTKG
	151	NAMAIAGARK MLESGASVGV LYCHSRRSVA QNYBVGCCQ HIGNFGAEYL
	201	ERRKQRYFEQ EGKLGFNSQS GKWERDQFKS YWKTKWYQDR DAPOELQKYI
	251	EGRHDKSWREN LAPQYDITPI DPSSLKQGS AGLFLKLEYQDG VFNKTYAQFR
10	301	DLNTKIGSRK IINRNQYQFN GLSINAPYTNL NLTAAYNNSGR QKYPKGSKFT
	351	GWGLKKDFET YNNAKLLDLX NTSTFRLPRE TELQTTLGPN YFHNEYGKNR
	401	FPEELGLFPD GDXDNGLYS YLGRFKGDKG LLLPKQSTIVQ PAGSQYFNTF
	451	YFDDAALKKDIYR YRLNYSNTV GYRFQGXYTS YYXSDEDFKR AFGESENPTYX
	501	KHCNMQSCGYI EFPVLRKRYGK RANNHSVIS ADFFGDYMF P ASYRTHRMP
	551	NTQEMYMFPSQI GDSCVHTALK PERANTWQFC FNTYKKGLLK QDDILGLKLV
15	601	GYRSRIDIYX HNNVKGWDL NGNIPWSWSS TGLAYTIQHR NFKDVKHHHG
	651	FELELNYDXY RFFTFLNYSY QKSTQPTNFS DASESPNNAS KEDQLKQGYG
	701	LSRVSALPRD YGRLEVRGTRW GLNKLTLGGA MRYFGKSIRA TAAERYIDXT
	751	NGXTXTSNFRQ LGKRSIXQTE TLARQPLIFD XYAAYPEKXX LIIFRAEVKNL
	801	FDRRYIOPLD AGNDAATQRY YSSFDPDKDK EEVTCNDNT LCNGKYGCTS
	851	KSVLTNFARG XTFLITMSYK *

20 ORF133a and ORF133-1 show 94.3% identity in 871 aa overlap:

		10	20	30	40
	orf133a.pep	KDKKVFTDARAVSTRQDIFK SXENLDNIVR XIPGAFTXQX			
	orf133-1	EAQIQVLEDVHVKA KAKRVPKDKKVFDTARAVSTRQDIFK SXENLDNIVR XIPGAFTXQX			
25		10 20 30 40 50 60	10 20 30 40 50 60	10 20 30 40 50 60	10 20 30 40 50 60
	orf133a.pep	50 60 70 80 90 100	SGXSVLMIRXKDSGFGRVN TMDVGDITXFYSTSTDAGRAGGSSQFGASVDSNFXAGLDVVK		
30	orf133-1	70 80 90 100 110 120	SGIVSLNIRGDSGFGRVN TMDVGDITQFYSTSTDAGRAGGSSQFGASVDSNFIA GLDVVK		
	orf133a.pep	110 120 130 140 150 160	GFSFGSGAGINSLAGSANL RLTKXVDVVQVNXYGLLKGLTGTNSTKGNAAMAI GARKWL		
35	orf133-1	130 140 150 160 170 180	GFSFGSGAGINSLAGSANL RLTVGDVVQVNNTYGLLKGLTGTNSTKGNAAMAI GARKWL		
	orf133a.pep	170 180 190 200 210 220	ESGA SAVGVVL YLGHSSRVA QN YRVGGGQ HIGNFGAEYL E LER RKRQRF E QEGGLKFKNSGK		
40	orf133-1	190 200 210 220 230 240	ESGA SAVGVVL YLGHSSRVA QN YRVGGGQ HIGNFGAEYL E LER RKRQRF V QEGGLKFKNSGK		
	orf133a.pep	230 240 250 260 270 280	WERDFQKSYWKT KWKYQDRPAQELQK YIEGHDKSWRENLAPQYDITP IDPSSLXKQSAGN		
45	orf133-1	250 260 270 280 290	WERDLQRQWQK YKPYKHN NN -QELQK YIEHDOKSWRENLAPQYDITP IDPSSLXKQSAGN		
	orf133a.pep	290 300 310 320 330 340	LFKLEYDG VFNKTAQFR DLNTKIGSKR KIINRNQYF NYGLS LNPFY TNL NLTAA YNSGRQK		
50	orf133-1	300 310 320 330 340 350	LFKLEYDG VFNKTAQFR DLNTKIGSKR KIINRNQYF NYGLS LNPFY TNL NLTAA YNSGRQK		
	orf133a.pep	350 360 370 380 390 400	YPKGSKFTGWLX KDFETYNNAKL DLDLX NTSTFRLP RETE LQTTLGPN Y FHNEYGKNR FP		
55	orf133-1	360 370 380 390 400 410	YPKGSKFTGWLX KDFETYNNAKL DLDL NNTAT FRLP RETE LQTTLGPN Y FHNEYGKNR FP		
	orf133a.pep	410 420 430 440 450 460	EELGLFFGDPDXDNGLYS YLGRFKGDKG LLLPKQSTIVQ PAGSQYFNTF YFDAA LKKDIYR		
60	orf133-1	420 430 440 450 460 470	EELGLFFGDPDQDNGLYS YLGRFKGDKG LLLPKQSTIVQ PAGSQYFNTF YFDAA LKKDIYR		

		470	480	490	500	510	520
	orf133a.pep	LNYSTNTVGYRPGGXYTGYYXXSDEFFKRAFGENSPTYXXKHCNQSCGIYEPVULKYQKRA					
5	orf133-1	LNYSNTNTVGYRPGGXYTGYYXXSDEFFKRAFGENSPTYXXKHCNQSCGIYEPVULKYQKRA					
		480	490	500	510	520	530
	orf133a.pep						
10	orf133-1	NNHSVSISADFGDYFMPFASYSRTHRMNPNIQEMYFSQIGDSGVHTALKPERANTWQFGFN					
		NNHSVSISADFGDYFMPFASYSRTHRMNPNIQEMYFSQIGDSGVHTALKPERANTWQFGFN					
		540	550	560	570	580	590
	orf133a.pep						
15	orf133-1	TYYKGLLQQDDILGLKLVLGYRSRISDIXYIHNVYGKWWDLNGN1PSWVSVSTGLAYTIQHNRF					
		TYYKGLLQQDDILGLKLVLGYRSRISDIXYIHNVYGKWWDLNGN1PSWVSVSTGLAYTIQHNRF					
		600	610	620	630	640	650
	orf133a.pep						
20	orf133-1	KDKVHKIGFELELNYDXXRFITNLISYAYQKSTQPTNFSDASESPPNASKEQDLQKQGYGLS					
		KDKVHKIGFELELNYDXXRFITNLISYAYQKSTQPTNFSDASESPPNASKEQDLQKQGYGLS					
		660	670	680	690	700	710
	orf133a.pep						
25	orf133-1	RVSALPRDYGRLEVGTWRLGNKLTLLGGAMRYFGKSIRATAEERYIDXTNGXXTSFRQLG					
		RVSALPRDYGRLEVGTWRLGNKLTLLGGAMRYFGKSIRATAEERYIDXTNGXXTSFRQLG					
		720	730	740	750	760	770
	orf133a.pep						
30	orf133-1	KRSIKQTETLARQPLIFDXAYAYPEPKXLIIFRAEVKNLFDRRYIDPLDAGNDATQRYY					
		KRSIKQTETLARQPLIFDXAYAYPEPKXLIIFRAEVKNLFDRRYIDPLDAGNDATQRYY					
		780	790	800	810	820	830
	orf133a.pep						
35	orf133-1	SFDPKDKDDEEVTCDNDTLCLNGKYGCCTSKSVLTFNARGXTFLITMSYKFX					
		SFDPKDKDDEEVTCDNDTLCLNGKYGCCTSKSVLTFNARGXTFLITMSYKFX					
		840	850	860	870	880	890
	orf133a.pep						
40	orf133-1						

Homology with a predicted ORF from *N.gonorrhoeae*

ORF133 shows 92.3% identity over 392 aa overlap with a predicted ORF (ORF133ng) from *N.*

45 gonorrhoeae:

	orf133.pep	PGYYCSDDDEFKRAFGCNSPTXKKHCNRSQCI	31
	orf133ng	FYFDAALKKDIFYRLNYSITNAINYRFGCEYTGYYCSENEFKRAFGCNSPAYKEHCDPSGCL	560
50	orf133.pep	YEPVLUKKYGGKRRANNNHSVSISADFGDYFMPFASYSRTHRMNPNIQEMYFSQIGDSGVHTAL	91
	orf133ng	YEPVLUKKYGGKRRANNNHSVSISADFGDYFMPFASYSRTHRMNPNIQEMYFSQIGDSGVHTAL	620
	orf133.pep	KFERANTWQPGFXTYKKGLLKQDDTLGLKLVLGYRSRISDNYIHNVYGKWWDLNGD1PSWVS	151
55	orf133ng	KFERANTWQPGFTYKKGLLKQDDTLGLKLVLGYRSRISDNYIHNVYGKWWDLNGD1PSWVS	680
	orf133.pep	STGLAYTIQHRXFDXDVHQXXXXXXYYDGRFTNLISYAYQKSTQPTNFSDASESPPNA	211
60	orf133ng	STGLAYTIQHRXFDXDVHQXXXXXXYYDGRFTNLISYAYQKSTQPTNFSDASESPPNA	740
	orf133.pep	SKEDOLQKQGYGLSRVSLPRDYGRLEVGTWRLGNKLTLLGGAMRYFGKSIRATAEERYIDG	271
	orf133ng	SKEDOLQKQGYGLSRVSLPRDYGRLEVGTWRLGNKLTLLGGAMRYFGKSIRATAEERYIDG	800
65	orf133.pep	TNGGNTSNFQLGKHSIKQTETLARQPLIFDXAYAYPEPKXLIIFRAEVKNLFDRRYIDPL	331
	orf133ng	TNGGNTSNFQLGKHSIKQTETLARQPLIFDXAYAYPEPKXLIIFRAEVKNLFDRRYIDPL	860

	orf133.pep	DAGNDAAXERYYSSFDPKDKDXDVTCAKDTCNGKYGGSKSVLNTNFARGRTFLMTMSY	391
	orf133ng	DAGNDAATQRYYSSFPDKDKEVDTCNAKDTCNGKYGGSKSVLNTNFARGRTFLMTMSY	920
5	orf133.pep	KF 393	
	orf133ng	KF 922	

The complete length ORF133ng nucleotide sequence <SEQ ID 881> is predicted to encode a
10 protein having amino acid sequence <SEQ ID 882>:

	1	MRSSFLRKPI CFYIMGVMLI HHSYAEDAGR AGSEAQIQVLF EDHVVKAKRVR
	51	PKDKVFTDAA RAVSTQDVF KSGENLNDNIV RSIFGAFTQQ DKSNSGIVSLN
15	101	IRGDGSGFRGV GINSILAGSN LRTLGVDDVV QGNNTYGLLL KGLTGTNNSTK
	151	VVXGSGFSSGRV KWLGEASAVG VLVGHSGRRGV AQNYRVGGGG QHIGNFGEEY
	201	GNAAMARAGAR KWLGEASAVG VLVGHSGRRGV AQNYRVGGGG QHIGNFGEEY
	251	LERRKQYQVW QEGGLKRNAG SGKWRDLQW QWXTKWKYKK YEDPQELQKY
	301	IEEHKDSWRE NLAQPDYITD IDPSGLKQQS AGNLLNLEYD GVFNVKTYAQF
	351	RDLNTRIGRS KIINRNYQFN YGLSLSPPNTS LNLTAAYNNSG RQKYPKGAKF
20	401	TWGGNLKKDFR TYNNAKILDL NNTATFRLPLP ETELQJTTLG F NYFHNEYGRK
	451	RFFPEGLGF DGPDPQDNGLY SYLNRFKGDDA GLLPQKSTIV OPAGSQYFNT
	501	FYFDAALKKD IYRLNYSNTNA IYRFQGEYETI YGYGSENFKF RAFGENSPAY
	551	KEHCDCSGLP YEPVLKXYYKG KHNRSVSI SADFGDYFMP FAGYSPRTHRM
	601	PNIQEMYFSQ IGDSGVETAL KPERANNTQF GFNTYKKGKL KQDDLGLKL
	651	VGTRSRIDNY IHVNRYGRWWD LNGDIEISWVG STGLAYTIRH RNFKDKDVHHK
25	701	GFELELNDY GRFTFTNLSY YQKSTQPTNFE SDASBSPNNAA SKEDQLKQY
	751	GLRSVSLPDR DYGRLEVGTR WLGNKLTLGG AMRIFYGSRIS ATAEEERYIDG
	801	TNGNTNSVNL QLGKRSIKQT ETIARQPLIF DFYAAYPEPKK NLIFRAEVKRN
	851	LFDRYRIDPL DAGNDAATQR QYSSFDPEKDK DEDVTCNAKD TLCNGKRYGGT
	901	SKSVLTNFAT GRTFLMTMSY KF*

30 A variant was also identified, being encoded by the gonococcal DNA sequence <SEQ ID 883>:

	1	ATGAGATCTT CTTCTCGGTT GAAGCCGATT TGTTTTTATC TTATGGGTGT
	51	TATGCTPATAT CATCATAGT ATGCCGAAAGA TGCAGGGCGC CGGGGAGCGG
35	101	AGGCGGAGAT ACAGGTTTG AGAAGATGTC ACGCTAACGGC GAAGCCGCTA
	151	CCGAAGACAA AAAAGGTGTT TACCGATGCG CCTGGCGTAT CGACCGGTca
	201	gGATGTGTC AACCTCCCGG AAAACCTCGA CRACATCGTA CCCACGATAC
	251	CGGTGCGTT TACACACCAA GATAAAAGCT CGGGCATITGT GTCTTTGAAT
	301	ATTGGGGCGT ACAGCGGGTT CGGGCGGCTC ATATGATGG TGACACGGAT
	351	CACGCAAGAC TTTATTCGCA CTTCATACCGA TGCGGGCAAGG CGACGGCGTT
40	401	CATCTTAATT CGTCGTCATCG CTGCACAGCA ATTATTTATGC CGGACTGGAT
	451	GTGCTCAAAT CGACGCTTCAG CGCTCGGGCA GGCTATCACAA CCTTTCGCGG
	501	CTGGCGGAAT CTGCGCGCTA TAGGCGTGGG TGACGTCGTT CAGGGCAATA
	551	ATACCTACGG CTCGCTCTA AAAGGCTGTA CGGGCACCAA TTCAACAAA
	601	GGTAATGGCA TGCGCGCGAT AGGTGCGGGG AAATGCGTGG AAAGCGGAGC
	651	GTCCTGCGGT GTGTCCTAACG GGCAACAGCG CGGGCGGGTG CGCCAAATT
45	701	ACCGCTGGG CGGGCGGGG CAGCACATC GAAATTTCGG TGAAGAATAT
	751	CTGGAAACGG CAAACAGCAG ATATTTTGTA CRAGAGGGTG TTGAAATAAT
	801	CAATGGCGGG AGCGAAAATGGGAAATGGGAAAGGGG TTGCAAAGG CAATACTGGA
	851	AAACAAAGTG GTATAATGGTA TACGAAGAAC CCCAAGAAACT GCAAAAAATAC
	901	ATCGAAAGAC ATGATAAAG CTGGCGGAA AACCTGGGC CGCAATACGA
	951	CATCACCCCT ATGTCATCGT CGGCGCTGTA CGACGACTGCG CGACGGCAATC
50	1001	TGTTTAATTG AGAACATGAC GGCGATTTCA ATAATACAC CGCCCAATT
	1051	CGCGATTTAA ACACCAAGAT CGGCAGCGGC AAAATCATCA ACCGCAATT
	1101	TCAATCTAAT TACGTTTTCT CTTTGACCC CTATACCAAC CTCATCTGA
	1151	CGCGACGCTA CAATTGGCGC AGGCAGAAAT ATCCGAAAGG CGCGAAGTT
	1201	ACAGGCTGGG GGCTTTTAAGA AGATTTGAA ACCTACACAA ACGGGAAAT
	1251	CCTGGACCTC AACACACCG CCACCTTCG GCTGCGCGC GAAACCGAGT
	1301	TGCAAAACCC TTGGGCGCTC AATTATTTCC ACAAGGAATA CGCGAAAC
	1351	CGCTTCTCTA AACGATTGG GCTGTTTTG GACGGTCTCG ATCAGGACA
	1401	CGGCGTTAT TCTCTTTGG GGCCTGTTAA GGGCGATAAA GGCGTGTGTC
60	1451	CTTCAAAATG ACCATTGTC CAACCGGCC CGACGCCATAA TTCAACACG
	1501	TTCTACTCGG ATGGCCGGCT AAAAAAGAC ATTACCGCT TAAACTACAG
	1551	CACCAATGCA ATCAACTACG GTTTCGGCGG CGAATATACG GCGTATTACG
	1601	GCTCGAAACCA CGAATTTAG CGGCAATTGCG GAGAAAACCT CGCGGCATAC
	1651	AAGGAAACATT GGACGGCCAG CTGGGGCGCT TATGAAACCC TATTGAAAAA
	1701	ATAGCGCAA AAGGCCGCCA ACAACCATTC GGTCAAGCATT AGTGCAGGACT
65	1751	TCGGCGATTAA TTTCATGGCG TTCGCCGGCT ATTCCGCGC ACACCGTATG

-480-

1801	CCCCACATCC	AAGRAATGTA	TTTTTCCCCAR	ATCGGGCAGT	CCGGCGTTCA
1851	CACCCCTTA	AAACCAAGAC	GCGCAACAC	TTGCCAATT	GCCTTCATAA
1901	CCTATAAAA	AGGATTTGTA	AAACAAAGATG	ATATATTAGG	ATTPGAAACTG
1951	GTCGGCTACC	GGAGCCGCA	TGACAACTAC	ATCCACAAACG	TTTACGGGAA
2001	ATGGTGGGAT	TTGAGCGGG	ATATCCCGAG	CTGGGTGGC	AGCACCGGGC
2051	TTGGCTCACAC	CATCCGCRCA	CGAACATTTC	AAGACAAGT	GCACAAACAC
2101	GGTTTGAGC	TGAGAGCTGA	TTACGATTAT	GGGGGTTTT	TCACCAACCT
2151	TCTTGTAGCC	TATCAAAGAA	GCACCCCAACC	GACCAATTTC	AGGGATGGGA
2201	GCGAAATGCC	CAACAAATGC	tccaaAGAAAG	ACCAACTCAA	ACAAGGTTAT
2251	GGGCTGAGCA	GGGTTCCTCG	CTTGGCGGGA	GATTACGGAC	GTTTGGAAAGT
2301	CGGTACGGC	TGGTGGCGA	AAACAACTGAC	TTTGGGGCG	GGCAtgcgCT
2351	ATTGCGCA	GAGACCGCC	GGACGGCTGA	AAGGAAAGCTA	TATCGACGCG
2401	ACCAACCGGG	GAATAACCG	CAATGTCGG	CAACTGGGCA	AGGGTTCCAT
2451	CAAAACAAAC	GGGGACCTTG	CCCGACAGCC	TTTGGATTTT	GATTTTTACG
2501	CGGCTTACGA	GGCGAAAGAA	AACTTATTAT	TCGGCGCGA	AGTCAAAAAC
2551	CTGTTGACCA	GGGCTTGT	CGATTCGGCTC	ATGTCGGGCA	ATGATGCGGC
2601	AACGCAACGGT	TATTCAGCT	CGTTCGGACCC	GAAGGAAAG	GAGGAAAGCG
2651	TAACGCTGTA	TGCTGTAA	ACGGTGTGCA	ACGGCAAAATA	CGGGCGGACCA
2701	AGCRAAAAGCG	TATIGACCCA	TTTCGCGACG	GGACGCAACT	TCTTGATGAC
2751	GATGAGCTAC	AAGGTTTAA			

This corresponds to the amino acid sequence <SEQ ID 884; ORF133ng-1>:

1	MRSSFLRKPI	CFYLMGVMLY	HHSYAEDAGE	AGSEQAQIVQL	EDVHVVKAKRV
51	PDKKVFDTA	RAVSTQRDF	KSGENLDNIV	RSIPIGAFTQO	DKKGSIVSLSN
101	IRGDGGRFRV	NTMDVGDIT	FYSTSTDAGE	AGGSSOFGAS	VDSNFIAGLD
151	VVKGSFGSSA	GINSLSAGSAN	LRTLGVDDVV	QNNNTYGLL	KGLTGTNSTK
201	GNAMAAIGAR	KWLESAGASVG	VLYGHSGRRV	QNYRVRVGGG	QHIGNFGEY
251	LERRKQFYEV	QEGGLKFNAG	SGKWERDLQK	QYWTRKWKYK	YEDPOELOKY
301	IIEHKDKSWRE	ILAPQYDITP	IDPSGLKQKQS	AGNLFLKLEYD	GVNKYTAQF
351	RDLNTRIGSE	KIINRNYQFN	YLGLSLNPYTN	LNLTAAYNNG	RQKYPKGAKF
401	TGWLGLLKDFE	TYNNAKLIDL	NNNTAFRLPHE	ETELQQTLLGF	NYFINNEYGKN
451	RFFEELGLFF	DGPQDONGL	SYLGRFKGDK	GLLPOKSTTV	QPAGSQSYNT
501	FYDAAALKKD	IYRLNYSNTA	YIRNFGGEYT	GYYGSENEFK	RAYGENSPAY
551	KEHCPDSCSQL	YEPUVLKYYKGK	KRANNHVSVI	SADFGDYFMP	FAGYSRTHRM
601	PNIQEMYFSQ	IGDSGVHTAL	KPERANTWQF	GNNTYKKGLL	KQDQILCLKL
651	VGYRSRIDN	IHNVYGRWWD	LNQDIPSWVG	STGLAYTIRH	RNFKDVKVHH
701	GFELLENLRYD	GRFTFTNSY	YQKSTQPTNF	SDASESPNNNA	SKEDQIKRGY
751	GLRSVSLAEP	DYGRLEVGTR	WLQNLKLTLG	AMRYFGKSIR	ATAEERYIDG
801	TNGNTSNVE	QLGKRSIKQT	ETLARQPLQE	DFYAAYPEPK	NLIIFRAEVKN
851	LFDRRYIDPL	DAGNDAAATOR	YQSSFDPKDE	DEDVTCAADK	TLNCNGKYGGT
901	SKSVLTNFAR	GRTFLMTMSY	KF*		

ORF133ng-1 and ORF133-1 show 96.2% identity in 889 aa overlap:

10	orf133ng-1.pep	SFRLKPCFYLGVMLY	HHSYAEDAGRAGSEQAQIVQL	LEDVHVVKAKRV/PDKKKVFTDARAV	
45	orf133-1			EAQIQVLEDVHVVKAKRV/PDKKKVFTDARAV	10 20 30
50	orf133ng-1.pep	STRQDVFKSGENLDNIVR	SPIGAFTQODKSSGIVSLSNIRGD	SGFGRVNTMVDGTTQFTYS	70 80 90 100 110 120
orf133-1	STRQDIFKSSGENLDNIVR	SPIGAFTQODKSSGIVSLSNIRGD	SGFGRVNTMVDGTTQFTYS	40 50 60 70 80 90	
55	orf133ng-1.pep	TSTDAGRAGGSQFGASV	DSENFIAGL DVKGSFSGSAGINSLAGS	ANLRTLGVDDVQGN	130 140 150 160 170 180
orf133-1	TSTDAGRAGGSQFGASV	DSENFIAGL DVKGSFSGSAGINSLAGS	ANLRTLGVDDVQGN	100 110 120 130 140 150	
60	orf133ng-1.pep	NTYGLLNLKLTGTNSTKG	NAMAAIGARKWLESGAS	AVGVLYGHSGRRV	190 200 210 220 230 240
orf133-1	NTYGLLNLKLTGTNSTKG	NAMAAIGARKWLESGAS	AVGVLYGHSGRRV	160 170 180 190 200 210	
65	orf133ng-1.pep	GPNFGEYELRKKQQYFV	QEGLLRFNAGSGK	WERDLQRQYWKTRWYKKYEDPQELQKYIEE	250 260 270 280 290 300

-481-

	orf133-1	GNGFQASBYLERRKQRQYFVQEGALKFNNSDGGKWERDLQRQQWYKPKYKNYYN-QELQKQYIEE 220 230 240 250 260
5	orf133ng-1.pep	310 320 330 340 350 360 HDKSWRRENLAPOQYDITFIDPSGLKQQSAGNLFKLEYDGVNKYTAQFRDLNTRIGCSRKII
	orf133-1	HDKSWRRENLAPOQYDITFIDPSGLKQQSAGNLFKLEYDGVNKYTAQFRDLNTRIGCSRKII 270 280 290 300 310 320
10	orf133ng-1.pep	370 380 390 400 410 420 NRNYQFNYQLSLNPYTINLNLTAAAYNSGRQKYPKGAKFTGWLGLLKDFTETYNNAKILDLNNNT
	orf133-1	NRNYQFNYQLSLNPYTINLNLTAAAYNSGRQKYPKGAKFTGWLGLLKDFTETYNNAKILDLNNNT 330 340 350 360 370 380
15	orf133ng-1.pep	430 440 450 460 470 480 ATFLRPLPRETELQTTLGFNFYHNFNEYGKNRPFEEGLLFFDGPDQDNGLYSYLGRFKKGDKGLL
	orf133-1	ATFLRPLPRETELQTTLGFNFYHNFNEYGKNRPFEEGLLFFDGPDQDNGLYSYLGRFKKGDKGLL 390 400 410 420 430 440
20	orf133ng-1.pep	490 500 510 520 530 540 PKQKSTIVQPGSQYFNTFYFDAAKKDIIYRLNLYSTNAINYRFGEETYGGYSENEFKRAF
	orf133-1	PKQKSTIVQPGSQYFNTFYFDAAKKDIIYRLNLYSTNAINYRFGEETYGGYSENEFKRAF 450 460 470 480 490 500
25	orf133ng-1.pep	550 560 570 580 590 600 GENSPTYKEHCDPSCGLYEPVLKKYKGKRANNHHSVISIADFGDYFMPFAGYSRTHRMPNI
	orf133-1	GENSPTYKHHCNRCSCIYEPVLKKYKGKRANNHHSVISIADFGDYFMPFAGYSRTHRMPNI 510 520 530 540 550 560
30	orf133ng-1.pep	610 620 630 640 650 660 QEMYFSQIGDSGVHTALKPERANTWQFGNTYKKGLLKQDDILGLKLVGYRSR1DNYIHN
	orf133-1	QEMYFSQIGDSGVHTALKPERANTWQFGNTYKKGLLKQDDILGLKLVGYRSR1DNYIHN 570 580 590 600 610 620
35	orf133ng-1.pep	670 680 690 700 710 720 VYGKWWDLINGDIPSWVGSTGLAYTIRRNFKDKVHKHGFELENYNDYGRFTTNLSAYQK
	orf133-1	VYGKWWDLINGDIPSWVGSTGLAYTIRRNFKDKVHKHGFELENYNDYGRFTTNLSAYQK 630 640 650 660 670 680
40	orf133ng-1.pep	730 740 750 760 770 780 STQPTNFSDASESPPNNASKEDQLKQGYGLRSVSLAPRDYGRLEVGVTRWLGNKLTLGGAMR
	orf133-1	STQPTNFSDASESPPNNASKEDQLKQGYGLRSVSLAPRDYGRLEVGVTRWLGNKLTLGGAMR 690 700 710 720 730 740
45	orf133ng-1.pep	790 800 810 820 830 840 YFGKSI RATEAEERYIDGTNGNTSNVRQLGKRSIKQTEYLARQPLIFDFYAAYPEPKKNI
	orf133-1	YFGKSI RATEAEERYIDGTNGNTSNVRQLGKRSIKQTEYLARQPLIFDFYAAYPEPKKNI 750 760 770 780 790 800
50	orf133ng-1.pep	850 860 870 880 890 900 FRAEVKNLFDRRYIDPLDAGDNAATQRYSSFEDPKDKDEDVTNCNAKDTLCNGKYGTSKS
	orf133-1	FRAEVKNLFDRRYIDPLDAGDNAATQRYSSFEDPKDKDEDVTNCNAKDTLCNGKYGTSKS 810 820 830 840 850 860
55	orf133ng-1.pep	910 920 VLTNTPARGRTFLMTMSYKFX
	orf133-1	VLTNTPARGRTFLMTMSYKFX 870 880
60	orf133ng-1.pep	

70 In addition, ORF133ng-1 is homologous to a TonB-dependent receptor in *H.influenzae*:

sp|P45114|Y1C17_HAEIN PROBABLE TGN8-DEPENDENT RECEPTOR H11217 PRECURSOR
 >gi|1075372|pi|G66410 transferin binding protein 1 precursor (tbpl) homolog -
 Haemophilus influenzae (strain Rd KW20) >gi|1574147 (U32801) transferrin binding
 protein 1 precursor (tbpl) [Haemophilus influenzae] Length = 913
 Score = 930 bits (2377), Expect = 0.0
 Identities = 476/921 (51%), Positives = 619/921 (66%), Gaps = 72/921 (7%)

 Query: 38 QVLEEVHVKAKWPKDKVFTDARAVSTRQVFKSGENIDNIVS1PIGAFTQODKSSGV 97
 + L + V K + DKK FT+A-A STAR+VFK + +D ++R+SIPIGAFTQODKSSGV 98
 Sbjct: 29 ETLGQIDVWEKV1SNNDKPFTEAKAKSTRVNFKETD1QVIR1PIGAFTQODKSSGV 88

 Query: 98 SLNIRGDSGFVGNTMDCITQFYSTSTDAGRAGGSQFSQGASVSNSFIAGLDVVKGSFS 157
 S+N+RG+++GRVNTMVG+D+D+D+ +D+G+GGSSQFGA++D+NFIAG+D K +FS
 Sbjct: 89 SVNIRGENGGLGRVNTMVGTVTDFYSTALDGGQSQQGSSQFGAAIDPNFIAQGVDNKNSNFs 148

 Query: 158 GSAGINSLAGSANSLRTLGVDVQXXXXXXXXXXXXXXXXXXXXAAGIRKWLSEGA 217
 G+GIN+LAGSAN RTLGIVDV+ M RKWL+G
 Sbjct: 149 GASGINALAGSANFRTLGWNVDIVTDKDPGII1KGMGTGSNATKSNSNFTMAAGRKWLDDNG 208

 Query: 218 SGVGVLYGHRSRQVAQYVRRGGCGHQHIGNFQGYEYLRRKQFOYEVBFGLKFLNAGSGKWRD 277
 WG VVY+G+S R V+Q+Y+R +GGG+ + + G+L + K+Y F + G N G W D
 Sbjct: 209 YGVVVYQSYQREVBQDYK1-GGGERLASLQLGDILAKEKEAYF-RNAYI1NP-EQOWTPD 265

 Query: 278 LQRQWIK-----TKWY-----KKYSDPQELOK---YIEE 303
 L + +W + +Y + KR + D + +OK IEE
 Sbjct: 266 LSKHKWCSNCNKPDYQKNGDCSYRIGSAAKTRREIQLLELTNGKPKRDIBKLQKGNDGIE 325

 Query: 304 HDKSWRENLAQPYDITPIDPSGLKQOSQAGNLFKLEYDGVNENKYTAQFRDINTRIGSRK 363
 DKS- N QY + P+I + L + S + K EY R L + +IGRSKI
 Sbjct: 326 TDLSFERNVSAIPEEGSLSQSRSHLHKFVYEGDDHHONLGAOLRDLNKSNSR 384

 Query: 364 NRNYQFNGLSLSNPYTNLNLTAAYNSGROKYPKGAKFTGWGLLKDFTETYNNAKILDLNNT 423
 NRNYQ NY + N Y + LNL AA+N G+ YF KG F GW + T N A I +D+NN+
 Sbjct: 385 NRNYQVNQNFFNNNSYLLNMLMAHINAIKTIYFKGFFGAGVADLKTRNVAVINNNS 444

 Query: 424 ATFLRPRETELQTTLGPNYFHNEYGKKNRFPSEELGLFDGDPDQDONGLYS--LGRFRGDKG 481
 TF LP+E +L+T+LGLNFTNEY NEY KNRFPSEEL LE+++ DL GLYS+ GR+ G K
 Sbjct: 445 HTFLLPKEIDLKTGLNFTNEYNEYKRNRFPESELSSLFYNDASHDQGLYHSRKRGRYSGCTK 504

 Query: 482 LLPOSTIVOPAGS0YQNFNTTFYDAA1LKKD1YRLNYS7NNA1NRYFGEYGTGYGSEMEFKR 541
 LLPQ+S I+QP+G Q F T YFD AL K IY LNYS N +Y F GEY GY
 Sbjct: 505 LLPORSVILQPSGKQKEKFTPLASLKG1HLYSVNTHYAFN6GEVYGY----- 555

 Query: 542 AFGENSPPAYKEHCDCPSQVYFEPVLUKKGKRNANHVSISADFQDFMPFAGYSRTHRP 601
 EN+ + +EP+ + K G K+A NH+ +SA+ D YFM+ +YRTHRP
 Sbjct: 556 ---ENTAGQO-----INEPLIHKSGHKKFAFNHSATLSAELS1YDFMFPTTYSRTHRP 604

 Query: 602 NIQEMMFYSQDGSVHTALKPERANTWQGQNTYKKG1KGLLKVLYGVRISDNLY 661
 NIQEM+PSQF++ +GV+TALKPER++T+0 GNTFYKKG1 QDD+LG+LWLGCVRS I NYI
 Sbjct: 605 NIQEMMFYSQVSNAGVN1TALKPEQSDTYQLGNTYKKGLEFTQDDLVGKLVGYRSFIKNY 664

 Query: 662 HNVYWDLNGDQVWGSTCQYFATLIRRHNFKDVHKGFELELYDYGRRFTNLSY 721
 HNVY WW +P+W S G YT H+N+K V K G ELE+NYD GRFF N+SYA
 Sbjct: 665 HNVYGVWW---RDGMP7WAESNGFKYTIJAHQNYK1PVKKSGVELEINYDMDGRFFANVSY 722

 Query: 722 QKSTOPNFSASESPNNASKEDOLQKGGYLRVSLSALPRDYGRLEVGRTRW1GNK1T1GGA 781
 Q++ QPTN++DAS PNNAS+ED LKQGYGLRSVS LP+DGRLE+GTRW K1T1G A
 Sbjct: 723 QRTOQNTYADASPRPNNASQDE1LKGQGYGLRSVSLPMDKGRYLEGTRWFDQ1L1T1GJA 782

 Query: 782 MRYFKKSIRATAEBERYIDTGTNGNTNSVROLGKRSIKOTETLARQLPLIFDFYAYEPKRN 841
 RY+GKS RAT EE YI+G+ + +R+ + +K+TE + +Q+I D + +YEP K+
 Sbjct: 783 ARYKGSKSIRATTEEYINGSNSR-FKRN1TLLRRENNYAVKTED1KQPI1LDLH8YEP1KD 841

 Query: 842 LIFRAEVKLNFDRRYIDPLDAGDNAATQYRSSYDFPDKDDEDWTCAKDTLCKNGYGGTS 901
 LI +AEV+NL D+RY+DPLDAGDNAA+QRYYSS + + C D + C GG+
 Sbjct: 842 LIKAEAVKLNFDRRYIDPLDAGDNAQDQYSS---NNSIECAQDSSAC---GGSD 892

 Query: 902 KSVLTNFARGRTFLMTMSYKF 922
 K+V+NL NFARGRT+*****YKF
 Sbjct: 893 KTLYNVFARGRTYLISLNK 913

The underlined motif in the gonococcal protein (also present in the meningococcal protein) is predicted to be an ATP/GTP-binding site motif A (P-loop), and the analysis suggests that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

5 Example 104

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 885>

```

1  ATGAACTGTA  TTTCACGGTA  CATCATCCGT  CAAATGGCGG  TTATGGCGGT
51  TTACGCCGTC  CTTCCTTCC  TCGCTTTGTA  CAGCTTTTTT  GAAATCCTGT
101  AGCAACCGG  CAACCTCGGC  AAAGGCAGTT  ACGGCATATG  GGAATGCTG
151  GGCTCACCGC  CCCTCAAAAT  GCCCGCCGCG  GCCTACGAAAC  TGATTCCTCT
201  CGCCGCTCTT  ATCGGGGAA  TGCTCTCCCT  CAGGCAGCTT  GCGGCCGCGA
251  CGGCAACTGAC  CGTCATCAA  GCGAGCGGCA  TGAGCACCAA  AAAGCTGCTG
301  TTGATCTGT  CGCAGTTCGG  TTTTATTTTT  GCTATTGCGA  CGCTGGCGCT
351  CGGGCAATGG  GTTGCSSCCA  CACTGAGCCA  AAAAGCCGAA  AACATCAAAG
401  CGCGCCGCAAT  CAACGGCAA  ATCAGACCCG  GCAATACCGG  CCTTGGCGTG
451  AAAGAAAAAA  ACAGCGTGT  CAATGTGGC  GAAATGTTG  CCGACCAT..
```

This corresponds to the amino acid sequence <SEQ ID 886; ORF112>:

```

1  MNLISRYIIR  QMAMVMAVYAL  LAFLALYSFF  EILYETGNLG  KGSYGIWEML
51  GYTAALKMPAR  AYELIFPLAVL  IGGLVSLSQL  AAGSELTVIK  ASGMSTKKLL
201  LILSQFGFIF  AIAITVALGEW  VAPFTLSQKAE  NIKAAAINGK  ISTGNTGLWL
151  KEKNSVINVR  EMLEDFH...  


```

Further work revealed further partial nucleotide sequence <SEQ ID 887>:

```

1  ATGAACTGTA  TTTCACGGTA  CATCATCCGT  CAAATGGCGG  TTATGGCGGT
51  TTACGCCGTC  CTTCCTTCC  TCGCTTTGTA  CAGCTTTTTT  GAAATCCTGT
101  AGCAACCGG  CAACCTCGGC  AAAGGCAGTT  ACGGCATATG  GGAATGCTG
151  gGCTACACCG  CCCUCAAAAT  GCCCGCCGCG  GCCTACGAAAC  TGATTCCTCT
201  CGCCGCTCTT  ATCGGGGAA  TGCTCTCCCT  CAGGCAGCTT  GCGGCCGCGA
251  GGCACACTGAC  CGTCATCAA  GCGAGCGGCA  TGAGCACCAA  AAAGCTGCTG
301  TTGATCTGT  CGCAGTTCGG  TTTTATTTTT  GCTATGCGA  CCTTGGCGCT
351  CGGGCAATGG  GTTGCSSCCA  CACTGAGCCA  AAAAGCCGAA  AACATCAAAG
401  CGCGCCGCAAT  CAACGGCAA  ATCAGACCCG  GCAATACCGG  CCTTGGCGTG
451  AAAGAAAAAA  ACAGCGTGT  CAATGTGGC  GAAATGTTG  CCGACCAT..
```

This corresponds to the amino acid sequence <SEQ ID 888; ORF112-1>:

```

1  MNLISRYIIR  QMAMVMAVYAL  LAFLALYSFF  EILYETGNLG  KGSYGIWEML
51  GYTAALKMPAR  AYELIFPLAVL  IGGLVSLSQL  AAGSELTVIK  ASGMSTKKLL
101  LILSQFGFIF  AIAITVALGEW  VAPFTLSQKAE  NIKAAAINGK  ISTGNTGLWL
151  KEKNSVINVR  EMLEDFH...  

201  LKLNIRRSTLGL  EDKVEVIAA  EENWPISVKL  NMDDVLLVLPF  DQMSVGELELT
251  YIRHQNNSO  NTRYIAIAWW  RKLVYPAAW  VNALVAFAF  PQTTRRGNMG
301  LKLFGGICXG  LLFLHLAGRLF  GFTSQL...
```

Computer analysis of this amino acid sequence predicts two transmembrane domains and gave the following results:

Homology with a predicted ORF from *N. meningitidis* (strain A)

ORF112 shows 96.4% identity over a 166aa overlap with an ORF (ORF112a) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
5	orf112.pep	MNLISRYIIRQMAVMAYVALAFLALYSFFEILYETGNLKGGSYGIWEMXGYTALKMPAR					
	orf112a	MNLISRYIIRQMAVMAYVALAFLALYSFFEILYETGNLKGGSYGIWEMXGYTALKMXAR	10	20	30	40	50
10	orf112.pep						60
	orf112a		70	80	90	100	110
15	orf112.pep	AYELIPMLPLAVLIGGLVLSQLAAGSELTVIKASGMSTKKLLLISLQFGFIFAIATVALGEW					120
	orf112a	AYEIMPLAVLIGGLVXSXSQLAAGSELXVIKASGMSTKKLLLISLQFGFIFAIATVALGEW	70	80	90	100	110
20	orf112.pep		130	140	150	160	
	orf112a	VAPTLISQKAENIKAANIKISTGNTGLWLKEKNNSIINVREMLPDHTLLGIKIWARNDK	130	140	150	160	170
25	orf112a		190	200	210	220	230
		ELAEAVEADSVALNSDGSWQLKNIRRSTLGEDKVEVSIAAEEXWPISVKRNLMDVLLVKP	190	200	210	220	230
							240

The ORF112a nucleotide sequence <SEQ ID 889> is:

25	1	ATGAACCTGA	TTTCAGCTTA	CATCATCGT	CAAATGGCGG	TATGGCGGT	
	51	TTACGGGTC	CTTCGCTTC	TGCGTTTGT	CGACTTTTT	GGAAATCTGT	
	101	ACGAAACCCG	CAACCTGGC	AAAGGCAT	ACGGCATTAG	GGAAATGNTG	
	151	GGNTTACCGG	CCCTCAAAAT	GNCCGCCCGC	GCCTACGAC	TGATGCCCT	
	201	CGCGTCTCTT	ATCGCGGCGAC	TGGCTCTCTNT	CGGCCAGCTT	GGCCGCCGCA	
30	251	GCGAACATGAA	CGTCATCAA	CCCCACCGCA	TGACGACCAA	AAACCTGCTG	
	301	TTGATTCCTGT	CGCAGTCGG	TTTATTTTT	GCTATTGCC	CGCTCGCCT	
	351	CGGGGAAATG	GTTCGGGCCA	CACTGAGGCC	AAAAGCGGAA	AAACATCAAAG	
	401	CGCGGGCCAT	CAACGGCAAA	ATCAGTACCG	GCAATACCGG	CTTTTGCTG	
35	451	AAAGAAAAAA	ACACGATTA	CAATGTGGC	GAATGTTGTC	CGGACCCATAC	
	501	CTCTGTTGGC	ATTAAATATT	GGGCCGCGCA	CGATAAAAAA	GAACCTGGCAG	
	551	AGCGCTGGG	AGCGCATTCC	GGCGTTTGG	ACAGCGACGG	CAGTTGGCAG	
	601	TTGAAACAAAC	TCGGCCCGCA	CACGGCTTGG	GRAGACAAAG	TCGAGGTCTC	
	651	TATGGGGCC	GAAGAAANT	GGCCGATTTC	CTCTAAACCC	AACTTGATGG	
40	701	ACGTATTGCT	CGTCAAACCC	GACCCAATGT	CGCTGGCGG	ACTGACCAC	
	751	TACATCGGCC	ACCTTCACCC	NNACAGCCAA	AAACCCCGAA	TCTACGCCAT	
	801	CGCATGGTGG	CGCAAAATTGG	TTTACCCCGC	CGCACGCTGG	CTGATGGCCC	
	851	TGCTGCGCTT	TGCTCTTACCC	CCGCACAAACCC	CCCGCCACGG	CAATATGGGC	
	901	TTAAATTCCT	TGCGGCCGAT	CTGCTCTCGGA	TTGCTGTCTC	ACCTTGCCCG	
45	951	NCGGCTCTTC	NGGTATTACCA	GCCAACTCTA	CGGCATATCCC	CCCTTCCTCG	
	1001	NCGGCCCACT	ACCTTACCAT	GGCTTCGGCT	TGCTGCGCGT	TTGCTGTATA	
	1051	CGCAACACGG	AAAACCGCTA				

This encodes a protein having the amino acid sequence <SEQ ID 890>:

50	1	MNLISRYIIRQMAVMAYVALAFLALYSFF	EILYETGNLG	KGSYGIWEMX
	51	GYTALKMPAR	AYELIPMLPLAVL	IGGLVLSXSQLAAGSELXVIKASGMSTKKLL
	101	LILSQQFGFIF	AIATVALGEW	VAPTLISQKAENIKAANIKISTGNTGLWL
	151	LKEKNNSIINV	ENMLPDHTLG	IKIWARNDK
	201	LNKNIRRSTLIG	EDKVEVSIAA	EEXWPISVKRNLMDVLLVKP
	251	VIRHLQXXSQ	NTRIYAIAWW	RKLVYPAAAW
	301	351	XFTPSOLYGIP	VMALVAFATT
55		RLKQEGK*	PFLXGALPTI	PCITRHGNMG

ORF112a and ORF112-1 show 96.3% identity in 326 aa overlap:

60	orf112a.pep	MNLISRYIIRQMAVMAYVALAFLALYSFFEILYETGNLGKGSYGIWEMXGYTALKMXAR		
	orf112-1	MNLISRYIIRQMAVMAYVALAFLALYSFFEILYETGNLGKGSYGIWEMXGYTALKPAR		
	orf112a.pep	AYELIPMLPLAVLIGGLVLSXSQLAAGSELXVIKASGMSTKKLLLISLQFGFIFAIATVALGEW		

	orf112-1	: : : : : AYELIPLAVLIGGLVLSQLAAGSELTIVIKASGMSTKKLLLILSQPGFIFAIATVALGEW
5	orf112a.pep	VAPTLSQKAENIKAAGINGKISTGNTGLWLKEKNSIINVREMLPDHTLLGIKIWARNDKN
	orf112-1	VAPTLSQKAENIKAAGINGKISTGNTGLWLKEKNSIINVREMLPDHTLLGIKIWARNDKN
	orf112a.pep	ELAEAVEADSAVLNSDGSWQLKNIRRLSTGEDKEVEVSIAAEEXWPISVKRNLMMDVLLVKP
10	orf112-1	ELAEAVEADSAVLNSDGSWQLKNIRRLSTGEDKEVEVSIAAEENWPISVKRNLMMDVLLVKP
	orf112a.pep	DQMSVGELETTYIRHLQXXSNQTRIYIAIWAWRKLVYPAAAAMVALVAFATTPOTRHGNMG
	orf112-1	DQMSVGELETTYIRHLQXXSNQTRIYIAIWAWRKLVYPAAAAMVALVAFATTPOTRHGNMG
15	orf112a.pep	LXKFGGICLGFLFHLAGRLXFXTSOLYGIPPFIXLGALPTIAFALLAVWLKRQEKRX
	orf112-1	LXKFGGICGXGLLFLFHLAGRLXFXTSOLYGIPPFIXLGALPTIAFALLAVWLKRQEKRX
20	Homology with a predicted ORF from <i>N.gonorrhoeae</i>	
	ORF112 shows 95.8% identity over 166aa overlap with a predicted ORF (ORF112ng) from <i>N. gonorrhoeae</i> :	
	orf112.pep	MNLISRYIIRQMAVMAVYALLAFLALYSFFEILYETGNLGKGSYGIWEMLGYTALKMPAR 60
25	orf112ng	MNLISRYIIRQMAVMAVYALLAFLALYSFFEILYETGNLGKGSYGIWEMLGYTALKMPAR 60
	orf112.pep	AYELIPLAVLIGGLVLSQLAAGSELTIVIKASGMSTKKLLLILSQPGFIFAIATVALGEW 120
	orf112ng	AYELIPLAVLIGGLVLSQLAAGSELTIVIKASGMSTKKLLLILSQPGFIFAIATVALGEW 120
30	orf112.pep	VAPTLSQKAENIKAAGINGKISTGNTGLWLKEKNSIINVREMLPDH 166
	orf112ng	VAPTLSQKAENIKAAGINGKISTGNTGLWLKEKTSIINVRGMLPDHTLLGIKIWARNDKN 180

The complete length ORF112ng nucleotide sequence <SEQ ID 891> is:

35	1	ATGAACCTGA	TTTACGCTTA	CATCATCCG	CAAATGGCGG	TTATGGCGGT
	51	TTACGGCGTC	CTTCGCTTC	TCCGTTTGT	CAGCCTTTTT	GAATTCCTGT
	101	ACGAACCCG	CAACCTGCC	AAAGCAGTC	ACGGCATATG	GGAAATGCTG
	151	GGCTACACCC	CCCTCACAACT	GCCGCCGCCG	GCCTACGAACT	TCATGCCCT
	201	CGCCGCTTC	ATCCGGCGAC	TGGCCCTCTC	CGCCGACTCT	GGCCGCCGCA
	251	GCGCACTGGC	COTCATCAAA	CCACAGCGCA	TGAGCACCCAA	AAAGCTGCTG
	301	TTGATTCCTGT	CTCACGTCG	TTTATTTTT	GCTATTGCCC	CCGTGCGCCT
	351	CGGCGGAATGG	GTTCGGGCCA	CGCTGAGGCC	AAAAGCGGAA	AACATCAAAag
	401	CCGGCGCCAT	taacggCAA	ATCAGCCacg	gcAAATACCGG	CCTT7tgCTG
	451	AAAGAAAAAA	ccACGATAT	TAATGTCG	GGAAATGTTGC	CCGACCCATAC
	501	GCTTTGGGC	ATCACAAATT	GGGGCGGCAA	CGATAAARAC	GAATTGCGAG
	551	AGGCAGTGG	AGCCGATTC	GGCTTTTGA	ACAGCGACGG	CAGCTGGCAG
	601	TTGAAAACAA	TCCGGCCAG	CATCATGGT	ACAGACAAAA	TGAAACATC
	651	cggCGCCGCC	GAAGAACAT	ggCGGATTTG	CCTCAGACGC	ACCTTGATGG
	701	ACGTATGGT	CGTCAGACCC	GACCAAATGT	CGCTGGCGA	GCTGACCAACC
	751	TACATCGCC	ACTCTCAA	CAACAGCCAA	AAACACCCAA	TCTACGCCAT
	801	CGCATGGTG	CGTAACACTG	TTTACCCCGT	CGCCGCATGG	GTCTATGGCC
	851	TCGTTGCTT	CGCCCTTACG	CGCACAACCA	CGGCCACCGG	CAATATGGGC
	901	TTAAACACTT	TGGCGGCAT	CTGTCCTCGA	TTGCTGTTC	ACCTTGGCGG
	951	CAGGCTCTTC	GGGTATTACCA	GCCAACCTCA	CGGCACCCCA	CCCTTCTCTG
	1001	CCGGCCGAC	GCCTACCCATA	GCCTTCGCT	TGCTCGTGT	TTGGCTGATA
	1051	CGCAACACGG	AAAAACCTTG	A		

This encodes a protein having amino acid sequence <SEQ ID 892>:

60	1	MNLISRYIIRQMAVMAVYALLAFLALYSFF	EILYETGNLGKGSYGIWEML
	51	GYTALKMPAR	A耶ELIPLAVLIGGLVLSQLAAGSELTIVIKASGMSTKKLL
	101	LILSQQGFI	FIIAAAVALEW VAPTLSQKAENIKAAGINGK
	151	KERTSISINVR	NIKAAAGINGKISTGNTGLWLKEKNSIINVREMLPDH
	201	LKNRIRSIM	ELAEAVEADSAVLNSDGSWQLKNIRRLSTGEDKEVEVSIAAE
	251	YIRHLQXXSNQ	EEWTPIAVRNLMDVLLVKP DQMSVGELETT
	301	LXKFGGICLG	LFLFHLAGRLXFXTSOLYGIPPFIXLGALPTIAFALLAVWLKR

351 RKQEKR*

ORF112ng and ORF112-1 show 94.2% identity in 326 aa overlap:

		10	20	30	40	50	60
5	orf112ng	MNLISRYIIRQMAVMAYVALLAFLALYSFFEILYETGNLGKGSYGIWEMLYTALKMPAR					
	orf112-1	MNLISRYIIRQMAVMAYVALLAFLALYSFFEILYETGNLGKGSYGIWEMLYTALKMPAR	10	20	30	40	50
10	orf112ng						
	orf112-1	AYELIPLAVLIGGLASLSQLAGSELVIKASGMSTKKLLLILSQFGFIPAIATVALGEN	70	80	90	100	110
15	orf112ng	AYELIPLAVLIGGLASLSQLAGSELVIKASGMSTKKLLLILSQFGFIPAIATVALGEN	70	80	90	100	110
	orf112-1	VAPTLSQKAENIKAIAINGKISTGNTGLWLKEKTSIIINVRGMLPDHTLLGIKIWARNDKN	130	140	150	160	170
20	orf112ng	VAPTLSQKAENIKAIAINGKISTGNTGLWLKEKTSIIINVRGMLPDHTLLGIKIWARNDKN	130	140	150	160	170
	orf112-1	ELAAEAVEADSVAVLNSDGSWQLKNIRRSIMGTDKIETSAAEETWPPIAVRRNLMMDVLLVKP	190	200	210	220	230
25	orf112ng	ELAAEAVEADSVAVLNSDGSWQLKNIRRSIMGTDKIETSAAEETWPPIAVRRNLMMDVLLVKP	190	200	210	220	230
	orf112-1	DQMSVGELETYYIRHLQNNSQNTQIYAIANWRKLVYPAAWVMALVAFAPTPOQTRHGNMG	250	260	270	280	290
30	orf112-1	DQMSVGELETYYIRHLQNNSQNTQIYAIANWRKLVYPAAWVMALVAFAPTPOQTRHGNMG	250	260	270	280	290
	orf112ng	LKLFGGICLGGLFHLAGRLPGFTSQLYGTGPFLAGALPTIAFALLAVWLIRKQEKRX	310	320	330	340	350
35	orf112-1	LKLFGGICLGGLFHLAGRLPGFTSQL	310	320			
	orf112ng						

This analysis suggests that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their

40 epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

It will be appreciated that the invention has been described by means of example only, and that modifications may be made whilst remaining within the spirit and scope of the invention.

TABLE I – PCR primers

ORF	Primer	Sequence	Restriction sites
ORF 1	Forward	CGCGGATCCGGTAGC-GGACACACTTATTCGG	BamHI-NheI
	Reverse	CCCG <u>CTCGAG</u> -CCAGCGGTAGCCTAATT XhoI	
ORF 2	Forward	GCGGATCCCATA <u>TATG</u> -TTTGATTTCGGTTGGG	BamHI-NdeI
	Reverse	CCCG <u>CTCGAG</u> -GACGGCATAACGGCG XhoI	
ORF 2-1	Forward	GCGGATCCCATA <u>TATG</u> -TTTGATTTCGGTTGGG	BamHI-NdeI
	Reverse	CCCG <u>CTCGAG</u> -TGATTTACGGACGCCA XhoI	
ORF 4	Forward	GCGGATCCCATA <u>TATG</u> -TGC GGAGGTCAAAAAGAC	BamHI-NdeI
	Reverse	CCCG <u>CTCGAG</u> -TTGGCTGGCCCTTC XhoI	
ORF 5	Forward	GGAATTCCATATGGCCATGG-TGGAAAGGCGCACAAACC	NdeI-NcoI
	Forward	CGGGATCC-ATGGAAAGGCCACAC	BamHI
	Reverse	CCCG <u>CTCGAG</u> -GACTGTGCAAAAACCGG XhoI	
ORF 6	Forward	CGCGGATCCCATA <u>TATG</u> -ACCCGTCAA <u>TCTCTGCA</u>	BamHI-NdeI
	Reverse	CCCG <u>CTCGAG</u> -TGC GCGGAACACTTTC XhoI	
ORF 7	Forward	CGCGGATCCGGTAGC-GCGCTGCTTTGGTCC	BamHI-NheI
	Reverse	CCCG <u>CTCGAG</u> -TTTCAAATATATTGCGGA XhoI	
ORF 8	Forward	GCGGATCCCATA <u>TATG</u> -GCTCAACTGCTTCGTAC	BamHI-NdeI
	Reverse	CCCG <u>CTCGAG</u> -AGCAGGCTTGGCGC XhoI	
ORF 9	Forward	CGCGGATCCCATA <u>TATG</u> -CCGAAGGAAGTCGGAAA	BamHI-NdeI
	Reverse	CCCG <u>CTCGAG</u> -TTTCGAGGGTTTCGGG XhoI	
ORF 10	Forward	GCGGATCCCATA <u>TATG</u> -GACACAAA <u>AGRAATCCTC</u>	BamHI-NdeI
	Reverse	CCCG <u>CTCGAG</u> -TAATGGGAA <u>ACCTTGT</u> TTT XhoI	
ORF 11	Forward	GCGGATCCCATA <u>TATG</u> -GCGGTCAACCTCTACG	BamHI-NdeI
	Reverse	CCCG <u>CTCGAG</u> -GGAAACGACTTCGCC XhoI	
ORF 13	Forward	CGCGGATCCCATA <u>TATG</u> -GCTCTGCTTCCGCGC	BamHI-NdeI
	Reverse	CCCG <u>CTCGAG</u> -AGGGTGTGTGATAATAAG XhoI	
ORF 15	Forward	GGAATTCCATATGGCCATGG-GCGGGACACTGACAG	NdeI-NcoI
	Forward	CGGGATCC-TGGGGACACTGACAGG	BamHI
	Reverse	CCCG <u>CTCGAG</u> -AGGTTGGCCTTGTCTATG XhoI	
ORF 17	Forward	GGAATTCCATATGGCCATGG -TTGCGGCCCTGTTCG	NdeI-NcoI

	Forward	CGGGATCC-ATTGCCGCCGTTCG	BamHI
	Reverse	CCCG <u>CTCGAG</u> -AAGCAGGTTGTACAGC	XhoI
ORF 18	Forward	CGGGATCCCAT <u>TATG</u> -ATTTGCTGCATTGGAT	BamHI-NdeI
	Reverse	CCCG <u>CTCGAG</u> -TCTTCCATTGGAAAGC	XhoI
ORF 19	Forward	GGAA <u>TTCCATATGGCCATGG</u> -TCGCCAGTGTTTACCC	NdeI-NcoI
	Forward	CGGGATCC-TTCGCCAGTGTTTACCC	BamHI
	Reverse	CCCG <u>CTCGAG</u> -GGTTTTTGAAAGCTGCC	XhoI
ORF 20	Forward	GGAA <u>TTCCATATGGCCATGG</u> -TCGGCGGGTATG	NdeI-NcoI
	Forward	CGGGATCC-TTCGGCGGGTATG	BamHI
	Reverse	CCCG <u>CTCGAG</u> -CGCGGAGCGAGCAC	XhoI
ORF 22	Forward	GGAA <u>TTCCATATGGCCATGG</u> -TGATTAATCAAAAAAGGTCT	NdeI-NcoI
	Forward	CGGGATCC-ATGATTAATCAAAAAAGGTCTAAACC	BamHI
	Reverse	CCCG <u>CTCGAG</u> -ATTATGATAGGCC	XhoI
ORF 23	Forward	CGCGATCCC <u>ATATG</u> -GATGTTCTGTTTAGAC	BamHI-NdeI
	Reverse	CCCG <u>CTCGAG</u> -TTTAAACCGATAGGTAACCG	XhoI
ORF 24	Forward	GGAA <u>TTCCATATGGCCATGG</u> -TGATGCCGAAATGGTG	NdeI-NcoI
	Forward	CGGGATCC-ATGATGCCGAAATGGTG	BamHI
	Reverse	CCCG <u>CTCGAG</u> -TGTAGCGGTGCCA	XhoI
ORF 25	Forward	CGGGATCCC <u>ATATG</u> -TATCGCAA <u>CTGATTGC</u>	BamHI-NdeI
	Reverse	CCCG <u>CTCGAG</u> -ATCGATGGAATAGCCG	XhoI
ORF 26	Forward	CGGGATCCC <u>ATATG</u> -CAGCTGATCGACTATT	BamHI-NdeI
	Reverse	CCCG <u>CTCGAG</u> -GACATCGGCCGTTT	XhoI
ORF 27	Forward	GGAA <u>TTCCATATGGCCATGG</u> -AGACCTATTCTGTTTA	NdeI-NcoI
	Forward	CGGGATCC-CAGACCTATTCTGTTTATTAACTC	BamHI
	Reverse	CCCG <u>CTCGAG</u> -GGGTTCGATTAARTAACCAT	XhoI
ORF 28	Forward	GGAA <u>TTCCATATGGCCATGG</u> -ACGGCTGTACGTTGATGT	NdeI-NcoI
	Forward	CGGGATCC-AACGGCTGTACGTTGATG	BamHI
	Reverse	CCCG <u>CTCGAG</u> -TTGTCAGAGGAATTGCC	XhoI
ORF 29	Forward	CGGGATCCC <u>ATATG</u> -AACGGTTGGATGCCG	BamHI-NdeI
	Forward	CGCGAT <u>CCCGCTAGC</u> -AACGGTTGGATGCCG	BamHI-NheI
	Reverse	CCCG <u>CTCGAG</u> -TTGCTCTAAGTCTCTGATATG	XhoI
ORF 32	Forward	CGCGATCCC <u>ATATG</u> -AATACTCCTCTTTG	BamHI-NdeI
	Reverse	CCCG <u>CTCGAG</u> -GCGTATTTTGATGCTTTG	XhoI
ORF 33	Forward	CGGGATCCC <u>ATATG</u> -ATTGATAGGATCGTATG	BamHI-NdeI
	Reverse	CCCG <u>CTCGAG</u> -TTGATCTTCAAACGCC	XhoI

ORF 35	Forward	GCGGATCCCATATG-TTCAGAGCTCAGCTT	BamHI-NdeI
	Forward	CGGGATCCGCTAGC-TTCAGAGCTCAGCTT	BamHI-NheI
	Reverse	CCCG <u>CTCGAG</u> -AACAGCCATTGAGCGA	XbaI
ORF 37	Forward	GCGGATCCCATATG-GATGACGTATCGGATTT	BamHI-NdeI
	Reverse	CCCG <u>CTCGAG</u> -ATAGCCCGCTTCAGG	XbaI
ORF 58	Forward	CGGGATCCGCTAGC-TCCGAACGCCAGTGGAT	BamHI-NheI
	Reverse	CCCG <u>CTCGAG</u> -AGCATTGTCCAAGGGGAC	XbaI
ORF 65	Forward	GGAATTCCATATGCCATGG -TGCTGTATCTGAATCAAG	NdeI-NcoI
	Forward	CGGGATCC-TTGCTGTATCTGAATCAAGG	BamHI
	Reverse	CCCG <u>CTCGAG</u> -CCGCATCGGCAGACA	XbaI
ORF 66	Forward	GCGGATCCCATATG-TACGCATTACCGCCG	BamHI-NdeI
	Reverse	CCCG <u>CTCGAG</u> -TGGATTTGCAAGAGATGG	XbaI
ORF 72	Forward	CGGGATCCCATATG- AATGCAGTAAARATATCTGA	BamHI-NdeI
	Reverse	CCCG <u>CTCGAG</u> -GCCTGAGACCTTGCAA	XbaI
ORF 73	Forward	CGGGATCCCATATG-AGATTTTCGGTATCGG	BamHI-NdeI
	Reverse	CCCG <u>CTCGAG</u> -TTCATCTTTCATGTTCG	XbaI
ORF 75	Forward	CGGGATCCCATATG- TCTGTCTTCAAAACGCC	BamHI-NdeI
	Reverse	CCCG <u>CTCGAG</u> -TTGTTTTGCAAGACAG	XbaI
ORF 76	Forward	GATCAGCTAGCCATATG-AAACAGAAAAAAACCGC	NheI-NdeI
	Reverse	CGGGATCC-TTACGGTTGACACCGTT	BamHI
ORF 79	Forward	CGGGATCCCATATG-GTTTCCGCCGCC	BamHI-NdeI
	Reverse	CCCG <u>CTCGAG</u> -GTGCTGATGCGCTTCG	XbaI
ORF 83	Forward	CGGGATCCCATATG-AAAACCTGCTGCTGC	BamHI-NdeI
	Reverse	CCCG <u>CTCGAG</u> -GCCGCCCTTGCGGCC	XbaI
ORF 84	Forward	CGGGATCCCATATG-GCAGAGATCTGTTG	BamHI-NdeI
	Reverse	CCCG <u>CTCGAG</u> -GTTTGCCGATCCGACCA	XbaI
ORF 85	Forward	CGGGATCCCATATG- GCGGTTGGGCCGGA	BamHI-NdeI
	Reverse	CCCG <u>CTCGAG</u> -TCGGCGCGGCCG	XbaI
ORF 89	Forward	GGAATTCCATATGCCATGG-CCATACCTCTTATCA	NdeI-NcoI
	Forward	CGGGATCC-GCCATACCTCTTATCAGAG	BamHI
	Reverse	CCCG <u>CTCGAG</u> -TTTTTGGCATTAGAAAAGC	XbaI
ORF 97	Forward	CGGGATCCCATATG-CATCCTGCCAGCGAAC	BamHI-NdeI

	Reverse	CCCGCTCGAG-TTCGCCCTACGGTTTTTG	XbaI
ORF 98	Forward	GGGGATCCCCATATG-ACGGTAACCTGCGG	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTGTGTTCCGGCAAATC	XbaI
ORF 100	Forward	GCGGATCCCCATATG-TCGGGCATTTACACCG	BamHI-NdeI
	Reverse	CCCGCTCGAG-ACGGGTTTCGGCGGA	XbaI
ORF 101	Forward	GCGGATCCCCATATG-ATTATCAAAAGAAACCTC	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTTCCGGCTTCAATGT	XbaI
ORF 102	Forward	GCGGATCCCCATATG-GCAGGGCTGTTTACC	BamHI-NdeI
	Reverse	CCCGCTCGAG-AAACGGTTTGAAACACGAC	XbaI
ORF 103	Forward	GCGGATCCCCATATG-AACCACGACATCAC	BamHI-NdeI
	Reverse	CCCGCTCGAG-CAGCCACAGGACGGC	XbaI
ORF 104	Forward	GCGGATCCCCATATG-ACGTGGGAACGC	BamHI-NdeI
	Reverse	CCCGCTCGAG-GCGCGTTTGAAACGGC	XbaI
ORF 105	Forward	GCGGATCCCCATATG-ACCAAATTCAAACCCCTC	BamHI-NdeI
	Reverse	CCCGCTCGAG-TAAACGAATGCGCTCCAG	XbaI
ORF 106	Forward	GCGGATCCCCATATG-AGGATAACCGAGCGGC	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTTGTTCCCGATGATGTT	XbaI
ORF 109	Forward	GCGGATCCCCATATG-GAAGATTATATATAATACTCG	BamHI-NdeI
	Reverse	CCCGCTCGAG-ATCAGCTTCGARCCGAAG	XbaI
ORF110	Forward	AAAGAATTC-ATGAGTAAATCCCGTAGATCTCCC	EcoRI
	Reverse	AAACTGCGAG-GGAAAAACACATCCGCACTCTGCC	PstI
ORF111	Forward	AAAGAATTC-GCACCGCAAAGGCCAAAACCGCA	EcoRI
	Reverse	AAACTGCGAG-TCTGCGCTTTCGGCGAGGTGG	PstI
ORF113	Forward	AAAGAATTC-ATGAAACAAAACCTCTATCGTGTGATTTCAACCG	EcoRI
	Reverse	AAACTGCGAG-TTACGAATGCCGTCTGCTCGACCGTACTG	PstI
ORF115	Forward	AAAGAATTC-TTGCTTGTGCAACAGAAAAAGACGG	EcoRI
	Reverse	AAAAAATGCGAC-CTATTTTATGGGGCTTTGCTGTGTTGAAAGCTGCC	Sall
ORF119	Forward	AAAGAATTC-TACAACATGTATCAGGAAAACCAATACCG	EcoRI
	Reverse	AAACTGCGAG-TTATGAAAACAGGCCAGGGCGTTTGCC	PstI
ORF120	Forward	AAAGAATTC-GCAAGGCTACCCCAATCCGCGGTG	EcoRI
	Reverse	AAACTGCGAG-CGGTTGGCTGCTGGCCGTTGAT	PstI
ORF121	Forward	AAAGAATTC-GCCTTGGCTGGCTGGTTTCGC	EcoRI
	Reverse	AAACTGCGAG-TCATCCGCCACCCACCTCGGCCATCCATC	PstI

ORF122	Forward Reverse	AAAAAA <u>GTCGAC</u> -ATGTC <u>TTACCGC</u> CAAGCAG <u>TTCTCC</u> <u>AAACTGCAG</u> -TCAGGAACAC <u>AAA</u> CGATGAC <u>GAATATCCG</u> TATC	SalI PstI
ORF125	Forward Reverse	AAAGAA <u>TTTC</u> -CGCTG <u>TTTTGCGGCCGG</u> GTAT <u>AAACTGCAG</u> -CGCCG <u>TTCAAGACGAAAAGTCG</u>	EcoRI PstI
ORF126	Forward Reverse	AAAGA <u>ATTTC</u> -CGGGAA <u>ACGGT</u> CGAAG <u>AAACTGCAG</u> -TTAAT <u>CTTGCTTCG</u> AATATAC	EcoRI PstI
ORF127	Forward Reverse	AAAGA <u>ATTTC</u> -ATGACTG <u>ATACTGGGGTT</u> TACG <u>AAAAAA<u>GTCGAC</u></u> -CTTAAG <u>TAACTTG</u> CAG <u>TCCTTATC</u>	EcoRI SalI
ORF128	Forward Reverse	AAAGA <u>ATTTC</u> -ATGCA <u>AGCTG</u> TC <u>CGCTACAGGCC</u> <u>AAACTGCAG</u> -CT <u>ATTCG</u> CA <u>RTG</u> CC <u>CCCGCCGG</u> AT <u>GTTGAGCGAGGCC</u>	EcoRI PstI
ORF129	Forward Reverse	AAAGA <u>ATTTC</u> -ATGG <u>ATTTG</u> CT <u>TTTGACATTA</u> TAC <u>GAATAACCG</u> <u>AAACTGCAG</u> - <u>TTATTTTG</u> AT <u>GAAATT</u> TT <u>GGGGCGG</u>	EcoRI PstI
ORF130	Forward Reverse	AAAGA <u>ATTTC</u> -CG <u>ACTTCCC</u> AT <u>U</u> CT <u>TCGGT</u> CG <u>AAACTGCAG</u> -C <u>I</u> CCGG <u>ATC</u> CT <u>GTAAACGC</u> ATT	EcoRI PstI
ORF 131	Forward Reverse	<u>GCGGATCCC</u> AT <u>ATG</u> -GAA <u>ATTCGG</u> CA <u>ATAAAAT</u> <u>CCCGCTCGAG</u> -CC <u>AGCGGAC</u> CG <u>CGTC</u>	BamHI-NdeI XhoI
ORF 132	Forward Reverse	<u>GCGGATCCC</u> AT <u>ATG</u> -AA <u>AGAAGCGGG</u> TT <u>TG</u> <u>CCCGCTCGAG</u> -CC <u>AATCTG</u> CC <u>AGCCGT</u>	BamHI-NdeI XhoI
ORF 133	Forward Reverse	<u>GCGGATCCC</u> AT <u>ATG</u> -GA <u>AGATG</u> CG <u>AGGC</u> CG <u>CCCGCTCGAG</u> -AA <u>ACTTG</u> AT <u>GCTCATCGT</u>	BamHI-NdeI XhoI
ORF 134	Forward Reverse	<u>GCGGATCCC</u> AT <u>ATG</u> -T <u>CTGTG</u> CA <u>AGCAGT</u> TT <u>TG</u> <u>CCCGCTCGAG</u> -AT <u>CTGTG</u> CC <u>AAATGCG</u>	BamHI-NdeI XhoI
ORF 135	Forward Reverse	<u>GCGGATCCC</u> AT <u>ATG</u> -CG <u>GTC</u> GA <u>AAAAG</u> TT <u>TT</u> <u>CCCGCTCGAG</u> -AA <u>ATACCG</u> T <u>GAGGATG</u>	BamHI-NdeI XhoI
ORF 136	Forward Reverse	<u>GCGGATCCC</u> AT <u>ATG</u> -AT <u>GAGGGCG</u> T <u>TAGCC</u> <u>CCCGCTCGAG</u> -TT <u>CCGA</u> AT <u>TTGGAA</u> CT <u>TTT</u>	BamHI-NheI XhoI
ORF 137	Forward Reverse	<u>GCGGATCCC</u> AT <u>ATG</u> -GG <u>CACGGCGG</u> AA <u>ATA</u> <u>CCCGCTCGAG</u> -AT <u>AACGGT</u> T <u>ATGCC</u> CC	BamHI-NdeI XhoI
ORF 138	Forward Reverse	<u>GCGGATCCC</u> AT <u>ATG</u> -TT <u>CGTTACAA</u> TT <u>CAGGC</u> <u>CCCGCTCGAG</u> -C <u>GGCG</u> TT <u>TATAGCGG</u>	BamHI-NdeI XhoI
ORF 139	Forward Reverse	<u>GCGGATCCC</u> AT <u>ATG</u> -G <u>CTTTTTGGCGG</u> TA <u>ATG</u> <u>CCCGCTCGAG</u> -TA <u>ACGTTCCG</u> T <u>CGCTT</u>	BamHI-NdeI XhoI

ORF 140	Forward Reverse	GCGGATCCCATATG-TTGCCCCACAGGCAGC CCCG <u>CTCGAG</u> -GACGATGGCAAACAGC	BamHI-NdeI XhoI
ORF 141	Forward Reverse	GCGGATCCCATATG-CCGTCTGAAGCAGTCT CCCG <u>CTCGAG</u> -ATCTGTTGTTTAAATATT	BamHI-NdeI XhoI
ORF 142	Forward Reverse	GCGGATCCCATATG-GATAATTCTGGTAGTGAAG CCCG <u>CTCGAG</u> -AACGTATAGCCTACCT	BamHI-NdeI XhoI
ORF 143	Forward Reverse	GCGGATCCCATATG-GATACCGGTTGAACCT CCCG <u>CTCGAG</u> -AATGGCTTCGGCAATATG	BamHI-NdeI XhoI
ORF 144	Forward Reverse	GCGGATCCCATATG-ACCTTTTACAACGTTGC CCCG <u>CTCGAG</u> -AGATTGTTGTTGTTTCG	BamHI-NdeI XhoI
ORF 147	Forward Reverse	GCGGATCCCATATG-TCTGTCTTCAAACGGC CCCG <u>CTCGAG</u> -TTTGTGTTGCAAGACAG	BamHI-NdeI XhoI

NB:

- restriction sites are underlined
 - for ORFs 110-130, where the ORF itself carries an *EcoRI* site (eg. ORF122), a *SaII* site was used in the forward primer instead. Similarly, where the ORF carries a *PstI* site (eg. ORFs 115 and 127), a *SaII* site was used in the reverse primer.
- 5

TABLE II – Summary of cloning, expression and purification

ORF	PCR/cloning	His-fusion expression	GST-fusion expression	Purification
orf 1	+	+	+	His-fusion
orf 2	+	+	+	GST-fusion
orf 2.1	+	n.d.	+	GST-fusion
orf 4	+	+	+	His-fusion
orf 5	+	n.d.	+	GST-fusion
orf 6	+	+	+	GST-fusion
orf 7	+	+	+	GST-fusion
orf 8	+	n.d.	n.d.	
orf 9	+	+	+	GST-fusion
orf 10	+	n.d.	n.d.	
orf 11	+	n.d.	n.d.	
orf 13	+	n.d.	+	GST-fusion
orf 15	+	+	+	GST-fusion
orf 17	+	n.d.	n.d.	
orf 18	+	n.d.	n.d.	
orf 19	+	n.d.	n.d.	
orf 20	+	n.d.	n.d.	
orf 22	+	+	+	GST-fusion
orf 23	+	+	+	His-fusion
orf 24	+	n.d.	n.d.	
orf 25	+	+	+	His-fusion
orf 26	+	n.d.	n.d.	
orf 27	+	+	+	GST-fusion
orf 28	+	+	+	GST-fusion
orf 29	+	n.d.	n.d.	
orf 32	+	+	+	His-fusion
orf 33	+	n.d.	n.d.	
orf 35	+	n.d.	n.d.	
orf 37	+	+	+	GST-fusion
orf 58	+	n.d.	n.d.	
orf 65	+	n.d.	n.d.	
orf 66	+	n.d.	n.d.	
orf 72	+	+	n.d.	His-fusion
orf 73	+	n.d.	+	n.d.
orf 75	+	n.d.	n.d.	
orf 76	+	+	n.d.	His-fusion
orf 79	+	+	n.d.	His-fusion
orf 83	+	n.d.	+	n.d.
orf 84	+	n.d.	n.d.	

orf 85	+	n.d.	+	GST-fusion
orf 89	+	n.d.	+	GST-fusion
orf 97	+	+	+	GST-fusion
orf 98	+	n.d.	n.d.	
orf 100	+	n.d.	n.d.	
orf 101	+	n.d.	n.d.	
orf 102	+	n.d.	n.d.	
orf 103	+	n.d.	n.d.	
orf 104	+	n.d.	n.d.	
orf 105	+	n.d.	n.d.	
orf 106	+	+	+	His-fusion
orf 109	+	n.d.	n.d.	
orf 110	+	n.d.	n.d.	
orf 111	+	+	n.d.	His-fusion
orf 113	+	+	n.d.	His-fusion
orf 115	n.d.	n.d.	n.d.	
orf 119	+	+	n.d.	His-fusion
orf 120	+	+	n.d.	His-fusion
orf 121	+	n.d.	n.d.	
orf 122	+	+	n.d.	His-fusion
orf 125	+	+	n.d.	His-fusion
orf 126	+	+	n.d.	His-fusion
orf 127	+	+	n.d.	His-fusion
orf 128	+	n.d.	n.d.	
orf 129	+	+	n.d.	His-fusion
orf 130	+	n.d.	n.d.	
orf 131	+	+	+	n.d.
orf 132	+	+	+	His-fusion
orf 133	+	n.d.	+	GST-fusion
orf 134	+	n.d.	n.d.	
orf 135	+	n.d.	n.d.	
orf 136	+	n.d.	n.d.	
orf 137	+	n.d.	+	GST-fusion
orf 138	+	n.d.	+	GST-fusion
orf 139	+	n.d.	n.d.	
orf 140	+	n.d.	n.d.	
orf 141	+	n.d.	n.d.	
orf 142	+	n.d.	n.d.	
orf 143	+	n.d.	n.d.	
orf 144	+	n.d.	+	n.d.
orf 147	+	n.d.	n.d.	

CLAIMS

1. A protein comprising an amino acid sequence selected from the group consisting of SEQ IDs 2, 4, 6, and 8.
2. A nucleic acid molecule which encodes a protein according to claim 1.
- 5 3. A nucleic acid molecule according to claim 2, comprising a nucleotide sequence selected from the group consisting of SEQ IDs 1, 3, 5, and 7.
4. A protein comprising an amino acid sequence selected from the group consisting of SEQ IDs 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252, 254, 256, 258, 260, 262, 264, 266, 268, 270, 272, 274, 276, 278, 280, 282, 284, 286, 288, 290, 292, 294, 296, 298, 300, 302, 15 304, 306, 308, 310, 312, 314, 316, 318, 320, 322, 324, 326, 328, 330, 332, 334, 336, 338, 340, 342, 344, 346, 348, 350, 352, 354, 356, 358, 360, 362, 364, 366, 368, 370, 372, 374, 376, 378, 380, 382, 384, 386, 388, 390, 392, 394, 396, 398, 400, 402, 404, 406, 408, 410, 412, 414, 416, 418, 420, 422, 424, 426, 428, 430, 432, 434, 436, 438, 440, 442, 444, 446, 448, 450, 452, 454, 456, 458, 460, 462, 464, 466, 468, 470, 472, 474, 476, 478, 480, 482, 484, 486, 488, 490, 492, 494, 496, 498, 500, 502, 20 504, 506, 508, 510, 512, 514, 516, 518, 520, 522, 524, 526, 528, 530, 532, 534, 536, 538, 540, 542, 544, 546, 548, 550, 552, 554, 556, 558, 560, 562, 564, 566, 568, 570, 572, 574, 576, 578, 580, 582, 584, 586, 588, 590, 592, 594, 596, 598, 600, 602, 604, 606, 608, 610, 612, 614, 616, 618, 620, 622, 624, 626, 628, 630, 632, 634, 636, 638, 640, 642, 644, 646, 648, 650, 652, 654, 656, 658, 660, 662, 664, 666, 668, 670, 672, 674, 676, 678, 680, 682, 684, 686, 688, 690, 692, 694, 696, 698, 700, 702, 25 704, 706, 708, 710, 712, 714, 716, 718, 720, 722, 724, 726, 728, 730, 732, 734, 736, 738, 740, 742, 744, 746, 748, 750, 752, 754, 756, 758, 760, 762, 764, 766, 768, 770, 772, 774, 776, 778, 780, 782, 784, 786, 788, 790, 792, 794, 796, 798, 800, 802, 804, 806, 808, 810, 812, 814, 816, 818, 820, 822, 824, 826, 828, 830, 832, 834, 836, 838, 840, 842, 844, 846, 848, 850, 852, 854, 856, 858, 860, 862, 864, 866, 868, 870, 872, 874, 876, 878, 880, 882, 884, 886, 888, 890, & 892..
- 30 5. A protein having 50% or greater sequence identity to a protein according to claim 4.

6. A protein comprising a fragment of an amino acid sequence selected from the group consisting of SEQ IDs 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134,
5 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252, 254, 256, 258, 260, 262, 264, 266, 268, 270, 272, 274, 276, 278, 280, 282, 284, 286, 288, 290, 292, 294, 296, 298, 300, 302, 304, 306, 308, 310, 312, 314, 316, 318, 320, 322, 324, 326, 328, 330, 332, 334,
10 336, 338, 340, 342, 344, 346, 348, 350, 352, 354, 356, 358, 360, 362, 364, 366, 368, 370, 372, 374, 376, 378, 380, 382, 384, 386, 388, 390, 392, 394, 396, 398, 400, 402, 404, 406, 408, 410, 412, 414, 416, 418, 420, 422, 424, 426, 428, 430, 432, 434, 436, 438, 440, 442, 444, 446, 448, 450, 452, 454, 456, 458, 460, 462, 464, 466, 468, 470, 472, 474, 476, 478, 480, 482, 484, 486, 488, 490, 492, 494, 496, 498, 500, 502, 504, 506, 508, 510, 512, 514, 516, 518, 520, 522, 524, 526, 528, 530, 532, 534,
15 536, 538, 540, 542, 544, 546, 548, 550, 552, 554, 556, 558, 560, 562, 564, 566, 568, 570, 572, 574, 576, 578, 580, 582, 584, 586, 588, 590, 592, 594, 596, 598, 600, 602, 604, 606, 608, 610, 612, 614, 616, 618, 620, 622, 624, 626, 628, 630, 632, 634, 636, 638, 640, 642, 644, 646, 648, 650, 652, 654, 656, 658, 660, 662, 664, 666, 668, 670, 672, 674, 676, 678, 680, 682, 684, 686, 688, 690, 692, 694, 696, 698, 700, 702, 704, 706, 708, 710, 712, 714, 716, 718, 720, 722, 724, 726, 728, 730, 732, 734,
20 736, 738, 740, 742, 744, 746, 748, 750, 752, 754, 756, 758, 760, 762, 764, 766, 768, 770, 772, 774, 776, 778, 780, 782, 784, 786, 788, 790, 792, 794, 796, 798, 800, 802, 804, 806, 808, 810, 812, 814, 816, 818, 820, 822, 824, 826, 828, 830, 832, 834, 836, 838, 840, 842, 844, 846, 848, 850, 852, 854, 856, 858, 860, 862, 864, 866, 868, 870, 872, 874, 876, 878, 880, 882, 884, 886, 888, 890, & 892..

7. An antibody which binds to a protein according to any one of claims 4 to 6.
- 25 8. A nucleic acid molecule which encodes a protein according to any one of claims 4 to 6.
9. A nucleic acid molecule according to claim 8, comprising a nucleotide sequence selected from the group consisting of SEQ IDs 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129,
30 131, 133, 135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253, 255, 257, 259, 261, 263, 265, 267, 269, 271, 273, 275, 277, 279, 281, 283, 285, 287, 289,

291, 293, 295, 297, 299, 301, 303, 305, 307, 309, 311, 313, 315, 317, 319, 321, 323, 325, 327, 329,
331, 333, 335, 337, 339, 341, 343, 345, 347, 349, 351, 353, 355, 357, 359, 361, 363, 365, 367, 369,
371, 373, 375, 377, 379, 381, 383, 385, 387, 389, 391, 393, 395, 397, 399, 401, 403, 405, 407, 409,
411, 413, 415, 417, 419, 421, 423, 425, 427, 429, 431, 433, 435, 437, 439, 441, 443, 445, 447, 449,
5 451, 453, 455, 457, 459, 461, 463, 465, 467, 469, 471, 473, 475, 477, 479, 481, 483, 485, 487, 489,
491, 493, 495, 497, 499, 501, 503, 505, 507, 509, 511, 513, 515, 517, 519, 521, 523, 525, 527, 529,
531, 533, 535, 537, 539, 541, 543, 545, 547, 549, 551, 553, 555, 557, 559, 561, 563, 565, 567, 569,
571, 573, 575, 577, 579, 581, 583, 585, 587, 589, 591, 593, 595, 597, 599, 601, 603, 605, 607, 609,
611, 613, 615, 617, 619, 621, 623, 625, 627, 629, 631, 633, 635, 637, 639, 641, 643, 645, 647, 649,
10 651, 653, 655, 657, 659, 661, 663, 665, 667, 669, 671, 673, 675, 677, 679, 681, 683, 685, 687, 689,
691, 693, 695, 697, 699, 701, 703, 705, 707, 709, 711, 713, 715, 717, 719, 721, 723, 725, 727, 729,
731, 733, 735, 737, 739, 741, 743, 745, 747, 749, 751, 753, 755, 757, 759, 761, 763, 765, 767, 769,
771, 773, 775, 777, 779, 781, 783, 785, 787, 789, 791, 793, 795, 797, 799, 801, 803, 805, 807, 809,
811, 813, 815, 817, 819, 821, 823, 825, 827, 829, 831, 833, 835, 837, 839, 841, 843, 845, 847, 849,
15 851, 853, 855, 857, 859, 861, 863, 865, 867, 869, 871, 873, 875, 877, 879, 881, 883, 885, 887, 889,
& 891..

10. A nucleic acid molecule comprising a fragment of a nucleotide sequence selected from the group consisting of SEQ IDs 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39,
41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91,
20 93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133,
135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173,
175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213,
215, 217, 219, 221, 223, 225, 227, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253,
255, 257, 259, 261, 263, 265, 267, 269, 271, 273, 275, 277, 279, 281, 283, 285, 287, 289, 291, 293,
25 295, 297, 299, 301, 303, 305, 307, 309, 311, 313, 315, 317, 319, 321, 323, 325, 327, 329, 331, 333,
335, 337, 339, 341, 343, 345, 347, 349, 351, 353, 355, 357, 359, 361, 363, 365, 367, 369, 371, 373,
375, 377, 379, 381, 383, 385, 387, 389, 391, 393, 395, 397, 399, 401, 403, 405, 407, 409, 411, 413,
415, 417, 419, 421, 423, 425, 427, 429, 431, 433, 435, 437, 439, 441, 443, 445, 447, 449, 451, 453,
455, 457, 459, 461, 463, 465, 467, 469, 471, 473, 475, 477, 479, 481, 483, 485, 487, 489, 491, 493,
30 495, 497, 499, 501, 503, 505, 507, 509, 511, 513, 515, 517, 519, 521, 523, 525, 527, 529, 531, 533,
535, 537, 539, 541, 543, 545, 547, 549, 551, 553, 555, 557, 559, 561, 563, 565, 567, 569, 571, 573,
575, 577, 579, 581, 583, 585, 587, 589, 591, 593, 595, 597, 599, 601, 603, 605, 607, 609, 611, 613,
615, 617, 619, 621, 623, 625, 627, 629, 631, 633, 635, 637, 639, 641, 643, 645, 647, 649, 651, 653,
655, 657, 659, 661, 663, 665, 667, 669, 671, 673, 675, 677, 679, 681, 683, 685, 687, 689, 691, 693,
35 695, 697, 699, 701, 703, 705, 707, 709, 711, 713, 715, 717, 719, 721, 723, 725, 727, 729, 731, 733,

735, 737, 739, 741, 743, 745, 747, 749, 751, 753, 755, 757, 759, 761, 763, 765, 767, 769, 771, 773,
775, 777, 779, 781, 783, 785, 787, 789, 791, 793, 795, 797, 799, 801, 803, 805, 807, 809, 811, 813,
815, 817, 819, 821, 823, 825, 827, 829, 831, 833, 835, 837, 839, 841, 843, 845, 847, 849, 851, 853,
855, 857, 859, 861, 863, 865, 867, 869, 871, 873, 875, 877, 879, 881, 883, 885, 887, 889, & 891..

- 5 11. A nucleic acid molecule comprising a nucleotide sequence complementary to a nucleic acid
molecule according to any one of claims 8 to 10.
12. A nucleic acid molecule comprising a nucleotide sequences having 50% or greater sequence
identity to a nucleic acid molecule according to any one of claims 8-11.
13. A nucleic acid molecule which can hybridise to a nucleic acid molecule according to any
10 one of claims 8-12 under high stringency conditions.
14. A composition comprising a protein, a nucleic acid molecule, or an antibody according to
any preceding claim.
15. A composition according to claim 14 being a vaccine composition or a diagnostic
composition.
- 15 16. A composition according to claim 14 or claim 15 for use as a pharmaceutical.
17. The use of a composition according to claim 14 in the manufacture of a medicament for the
treatment or prevention of infection due to Neisserial bacteria.

1/24

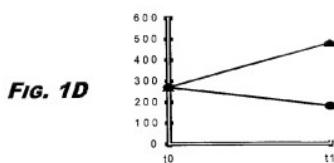
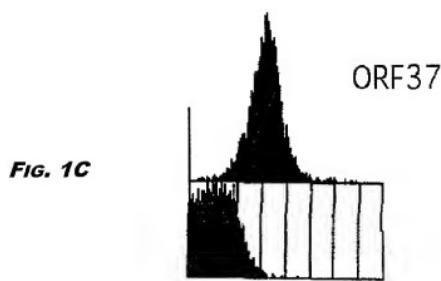
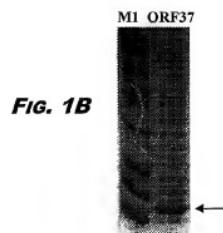
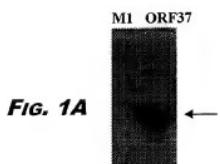
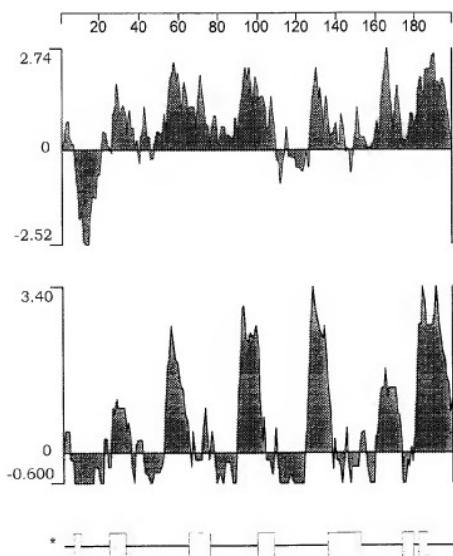
FIGURE 1

FIG 1E

3/24

FIGURE 2***FIG. 2A***

M1 ORF5

***FIG. 2B***

TP



FIGURE 3***FIG. 3A***

M1 ORF2

***FIG. 3B***

M1 ORF2

***FIG. 3C***

TP OMV

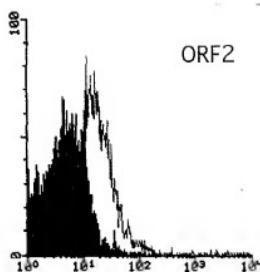
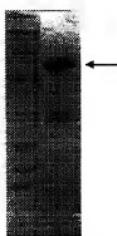
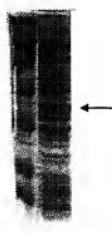
***FIG. 3D***

FIGURE 4***FIG. 4A***

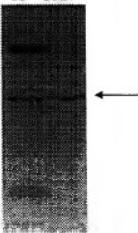
M1 ORF15

***FIG. 4B***

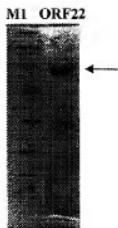
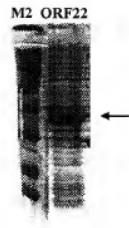
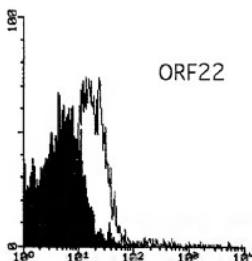
M2 ORF15

***FIG. 4C***

TP OMV



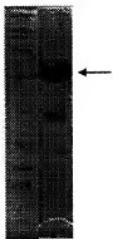
6/24

FIGURE 5**FIG. 5A****FIG. 5B****FIG. 5C**

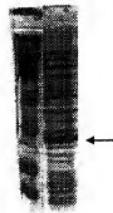
7/24

FIGURE 6***FIG. 6A***

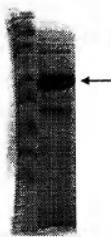
M1 ORF28

***FIG. 6B***

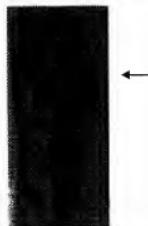
M2 ORF28

**FIGURE 7*****FIG. 7A***

M1 ORF32

***FIG. 7B***

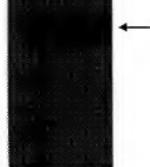
M1 ORF32



8/24

FIGURE 8

M1 ORF4

FIG. 8A

M1 ORF4

FIG. 8B

TP OMV

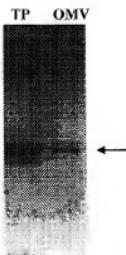
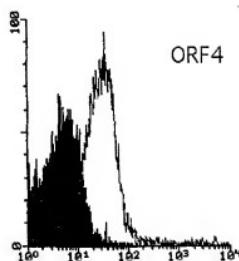
FIG. 8C**FIG. 8D**

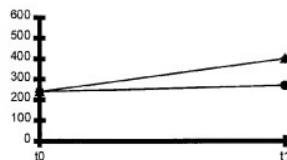
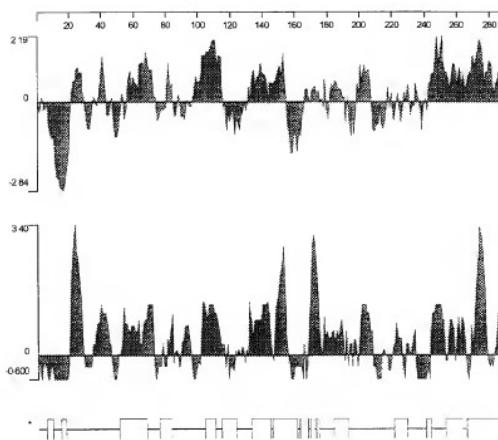
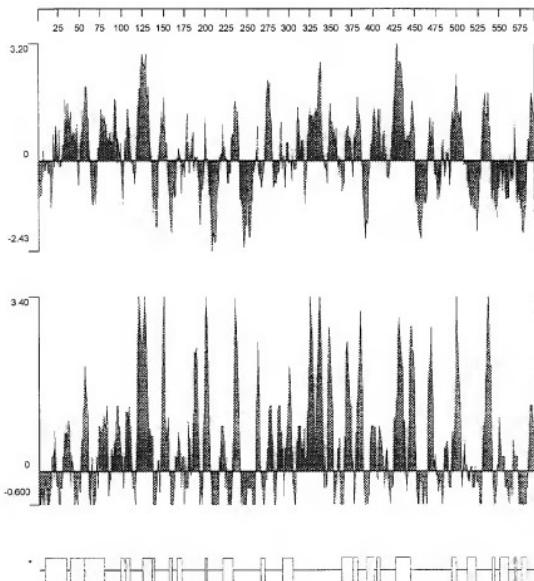
FIG. 8E**FIG. 8F**

FIGURE 9

11/24

FIGURE 10***FIG. 10A***

M1 ORF76

***FIG. 10B***

TP OMV

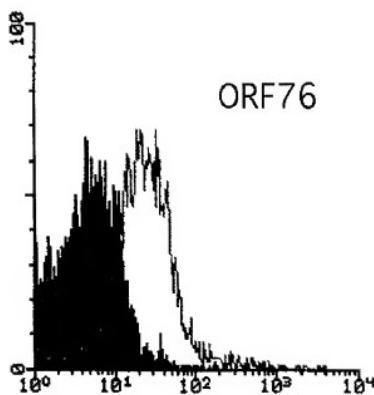
***FIG. 10C***

FIGURE 11

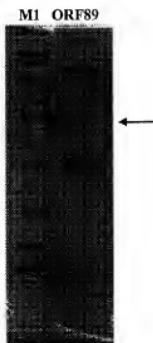
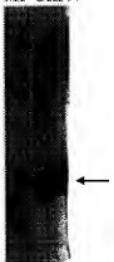
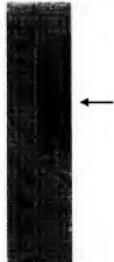


FIGURE 12***FIG. 12A***

M1 ORF97

***FIG. 12B***

M1 ORF97

***FIG. 12C***

TP OMV

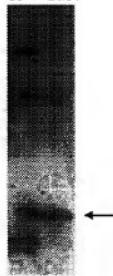
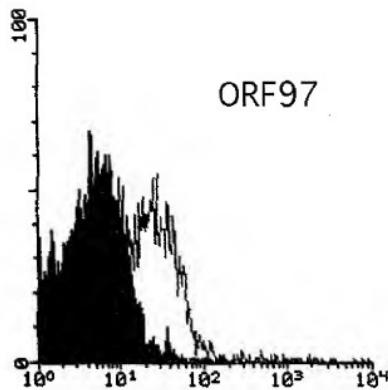
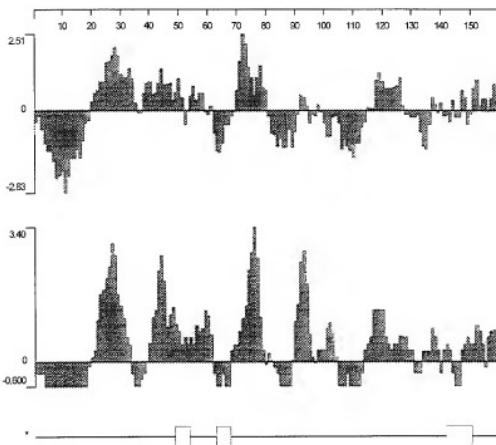
***FIG. 12D***

FIG. 12E

15/24

FIGURE 13***FIG. 13A***

M1 ORF106

***FIG. 13B***

M2 ORF106

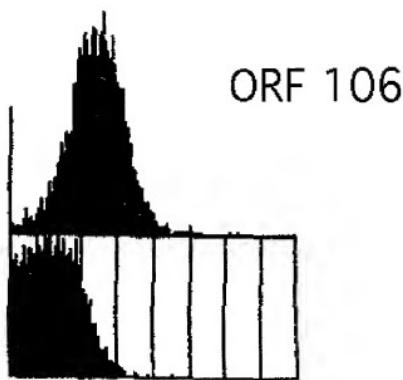
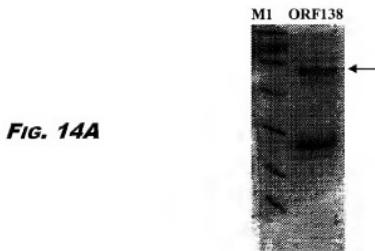
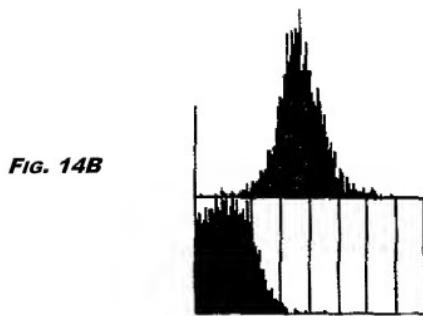
***FIG. 13C***

FIGURE 14**FIG. 14A****FIG. 14B**

17/24

FIGURE 15***FIG. 15A***

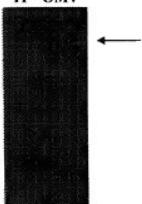
M1 ORF23

***FIG. 15B***

M2 ORF23

***FIG 15C***

TP OMV



18/24

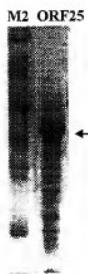
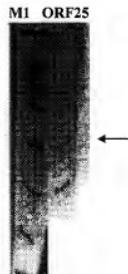
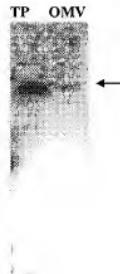
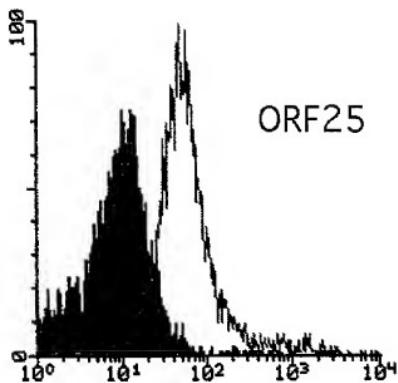
FIGURE 16**FIG. 16A****FIG. 16B****FIG. 16C****FIG. 16D**

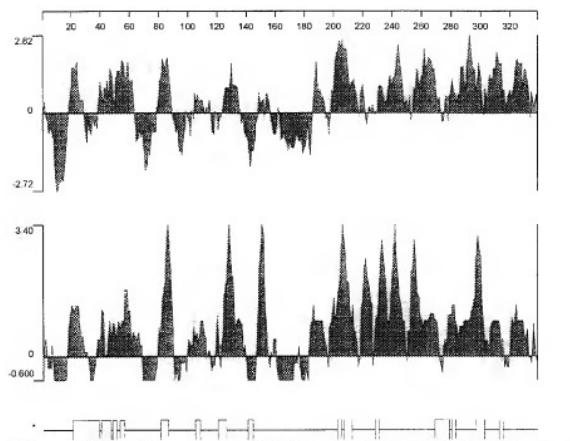
FIG. 16E

FIGURE 17***FIG. 17A***

M1 ORF27

***FIG. 17B***

M2 ORF27

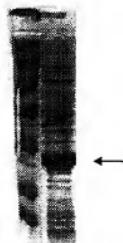
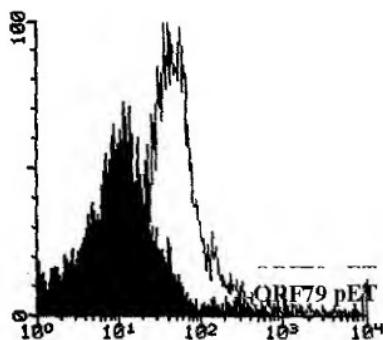


FIGURE 18***FIG. 18A******FIG. 18B***

22/24

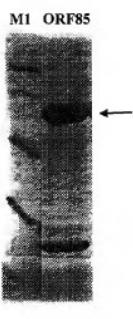
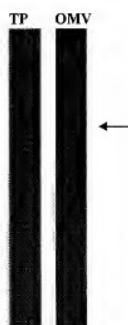
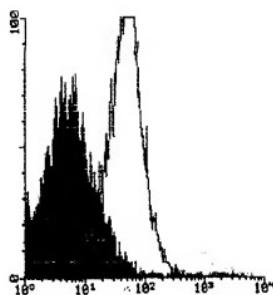
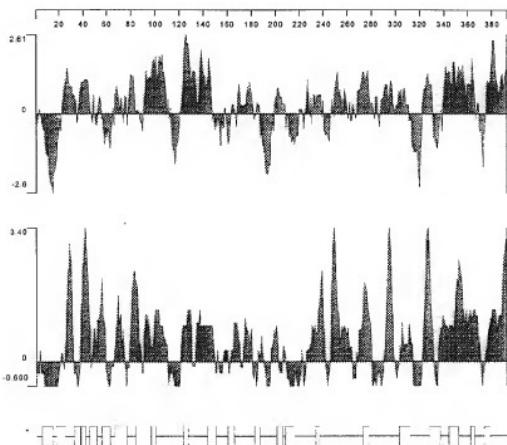
FIGURE 19***FIG. 19A******FIG. 19B******FIG. 19C***

FIG 19D



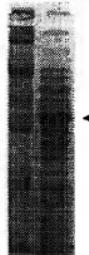
24/24

FIGURE 20***FIG. 20A***

M1 ORF132

***FIG. 20B***

M2 ORF132

***FIG. 20C***